REPORT RESUMES

ED 002 841 24

INVESTIGATION OF THE USE OF STATISTICS IN COUNSELING STUDENTS.

BY- HEWES, ROBERT F.
MASSACHUSETTS INST. OF TECH., CAMBRIDGE

REPORT NUMBER CRP-336 PUB DATE 1 JAN 60
EDRS PRICE MF-$0.18 HC-$3.80 95P.

DESCRIPTORS- *COLLEGE STUDENTS, *COUNSELING SERVICES, PROGRAM EVALUATION, *PROGRAM PLANNING, STUDENT DEVELOPMENT, *STUDENT PLACEMENT, MASSACHUSETTS, CAMBRIDGE

THE OBJECTIVE WAS TO EMPLOY TECHNIQUES OF PROFILE ANALYSIS TO DEVELOP THE JOINT PROBABILITY OF SELECTING A SUITABLE SUBJECT MAJOR AND OF ASSURING TO A HIGH DEGREE GRADUATION FROM COLLEGE WITH THAT MAJOR. THE SAMPLE INCLUDED 1,197 MIT FRESHMEN STUDENTS IN 1952-53, AND THE VALIDATION GROUP INCLUDED 699 ENTRANTS IN 1954. DATA INCLUDED SECONDARY SCHOOL GRADES, RANK AND TYPE OF SCHOOL ATTENDED, COLLEGE BOARD SCORES, INTENDED COLLEGE MAJOR AT TIME OF APPLICATION AND MATRICULATION, FIRST SEMESTER GRADES AT MIT (INDEPENDENT VARIABLES) GRADE AVERAGE AT MIT, COLLEGE MAJOR IN LAST SEMESTER OF ATTENDANCE, AND BACHELOR'S DEGREE EARNED (DEPENDENT VARIABLES). REGRESSION ANALYSIS WAS USED TO DERIVE A PREDICTION OF CHANCES FOR GRADUATION GIVEN MEMBERSHIP IN A PARTICULAR MAJOR. DISCRIMINANT ANALYSIS WAS EMPLOYED TO DERIVE THE PROBABILITY OF MEMBERSHIP IN A PARTICULAR MAJOR. THE CONCLUSIONS REACHED WERE THAT THE PROBABILITIES WOULD FAIL IN PASSING NEW AND RELEVANT INFORMATION TO THE STUDENT AS FAR AS HIS ACADEMIC SCORES ARE CONCERNED. (GC)

ED002841

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Investigation of the Use of Statistics in Counseling Students

Robert E. Hewes, Registrar
Richard W. Willard, Statistical Analyst

Massachusetts Institute of Technology
Cambridge, Massachusetts

January 1, 1960

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

This work was done in part at the M.I.T. Computation Center, Cambridge, Massachusetts.

3414

TABLE OF CONTENTS

	Page
Abstract	1
Counseling and Statistics	10
The Tatsuoka Procedures	21
Pilot Studies: Probability of Success	25
Clustering of Career Majors	45
Determination of variables to be Included in the Analyses	53
Computation of Prediction and Discriminant Functions	59
Tests with a New Group of Students - Class of 1958	69
Discussion of Results	83
Bibliography	88

TABLES

Table	Title	Page
I	Hypothetical IQ Distributions for Two Colleges	14
II	Hypothetical IQ Distributions for Two Colleges by Graduation and Failure	15
III	Variance-Covariance: 611 Graduates, Class of 1956	30
IV	Variance-Covariance: 211 Nongraduates, Class of 1956	31
V	Number Graduating at Various Deciles of Experimental Functions	35
VI	Average Grades in Freshman Subjects; Graduates; Nongraduates by College Major	41
VII	Cosines Between Some of the Difference Vectors from Table VI	42
VIII	Values of D^2 (Based on 8 M.I.T. Grades) Arranged in Increasing Order of Magnitude	48
IX	Inter-and Intra-Cluster Average Distances, D^2	52
X	Raw Regression Weights; Four Year Average as Criterion	60
XI	Correlations of Regression Function with Independent Variables	60
XII	Two Group Discriminant Function Weights; Graduates vs. Nongraduates as Criterion	62
XIII	Correlations of Graduates - Nongraduates Discriminant Functions with Independent Variables	62
XIV	Values of λ (Latent Roots) in Solutions to Matrix Equation 6.1	64
XV	Approximate Tests of Significance of Discriminant Functions	65
XVI	Weights for Three Basis Vectors of Reduced Space: Discrimination of Clusters of Majors: Chemistry group	66
XVII	Weights for Three Basis Vectors of Reduced Space: Discrimination of Clusters of Majors: Physics group	66

Tables (Continued)

Table	Title	Page
XVIII	Averages of Difference Between Probability of Given Cluster and "Perfect" Probability for Each Cluster	70
XIX	Student "t" Test Between Graduates and Nongraduates for Two Functions with Different Criteria	72
XX	Average Difference Between "Perfect" Probability of Graduation and Computed Probability: Class of 1958	78
XXI	Distribution of Regression Scores (Arbitrary Scale) for Class of 1958: Graduates, Nongraduates; Proportion in Each Interval Who Were Graduates; Chemistry Group	79
XXII	Distribution of Regression Scores (Arbitrary Scale) for Class of 1958: Graduates, Nongraduates; Proportion in Each Interval Who Were Graduates; Physics Group	79

FIGURES

Figure	Title	Page
1	Empirical Curves for Points Having 25%; 50%; and 75% of Students Graduating	27
2	Comparison of Logistic Curve and Proportions Graduating. Class of 1958: Chemistry Group	81
3	Comparison of Logistic Curve and Proportions Graduating. Class of 1958: Physics Group	82

ABSTRACT

This study was designed to investigate the potential use of certain statistics as an aid to the student faced with choosing his college major. On the premise that the choice of a college major involves consideration of both objective and subjective information, the intent was to summarize the objective information in a statistic which would facilitate the interaction of the two kinds of information in the student's decision.

A model for a student's choice of career would include any of several techniques under the category of profile analysis whereby the student's profile is compared to the profiles of students in the various majors. The model may include an evaluation of the differential risks involved in the various majors made in terms of likelihood of graduation. The model must also include some assessment of the student's interests and aspirations. The latter at present are best evaluated by the student himself possibly with the aid of an advisor. The first two considerations can be handled objectively. It is proposed to find a statistic which would combine both considerations and be of a form which enables its use by the student in his subjective treatment of all three considerations. By methods of multivariate analysis it is proposed that the joint probability of membership in a major and graduation in that major could be computed for each major available to the student. Such probabilities would serve to synthesize the profile analysis and the differential survival

predictions. Further, such probabilities should be understood by the individual student and be easily integrated with his interests and aspirations. The purpose of this study is to develop most effectively such probability statistics. The question of their actual success in the counseling of students who must choose a major is beyond this study and is virtually impossible to assess.

One strong logical reason for the use of such a probability statistic has to do with an individual's freedom of choice. By common techniques of profile analysis it is possible to identify the college major whose students have profiles most like the profile of the individual's profile. By common techniques of differential survival predictions it is possible to identify the college major in which the chances of graduation are the greatest. These two criteria might identify for the individual two different choices. In addition, his own preferences may involve a third choice. The question then is how to resolve all differences. The proposed statistic would synthesize the first two criteria. Its biggest advantage, however, lies in the fact than an individual could still select his own preference knowing in the form of risks and probabilities the extent to which he is making a choice contrary to the criteria of profile similarity and survival prediction.

In developing the proposed statistic analyses were made of 1197 students in the classes of 1956 and 1957 at M.I.T. Tests of the statistics were made on 699 students in the class of 1958. The choices of careers are made by M.I.T. students at the end of their freshman year. Many change their choices later, and their final choices were the ones used in this study. At the time of their choice in the freshman year the following significant academic measures are available:

1. Measures from admissions folders.
    a) High school grades in mathematics and science.
    b) Rank in high school graduating class.
    c) Scores in entrance tests of the CEEB.
    d) Rating of professional promise made by Admissions Office Staff.
2. Grades in first term at M.I.T.
    a) chemistry
    b) physics
    c) humanities
    d) mathematics

These served as the basis of the study.

Preliminary studies were made of the potential of non-linear models as the basis of the prediction of graduation. Considered were:

1. Generalized quadratic model
2. "Distance" function: geometric distance from point corresponding to maximum scores to point corresponding to given set of scores.
3. Quadratic form derived from parameters of the multivariate normal distribution function.
4. A generalized quadratic model involving transformed scores

None of these showed any improvement over the use of a simple linear function.

Another analysis was made of the comparison between two criteria for the derivation of the linear function serving as the basis of the prediction of success. One criterion involved the regression of four-year grade averages upon the independent measures, but this excluded those not reaching the fourth year. The other criterion used the group of graduates and the group of non-graduates as the criterion for a two group discriminant analysis. The latter included all but those who withdrew in good standing. There was no essential difference between the efficiency of the two systems.

The final analysis included comparisions between several kinds of probabilities. First, the probabilities of membership in a major were computed from individual profiles for members of the Class of 1958 using parameters estimated from members of the classes of 1956 and 1957. These were compared to _a priori_ probabilities, the proportions in the earlier classes choosing each major. This comparision showed that the analysis of profiles did not improve upon the rather trivial _a priori_ probabilities. Hence the profile analysis of the academic measures adds no useful information. The reason for this is the fact that individual variations within majors on the academic measures studied are greater in magnitude than any variations between majors. Stated in other terms, the overlapping of the various majors on these measures is too great to attach significance to the small differences between means.

Secondly, for each student in the class of 1958 two probabilities of graduation were computed from the academic measures using parameters derived from the earlier classes. One probability provided for differences between the majors in terms of difficulty of achieving graduation and used, besides the academic measures, the final choice of major. The other probability did not reflect differences between majors and did not use the choice of major. On comparison, the probability which reflected differences between majors was not any better than the other. This indicates that the differences between the various majors in difficulty as related to the measures studied is less than the differences that could be effectively used for differential prognoses. In short, a student with a poor prognosis in a more difficult major has so little better prognosis in a less difficult major that the "noise" inherent in the prediction system obscures the difference between the two prognoses. It must be remembered that this conclusion involves only the measures studied and excludes such factors as personal devotion (or lack of it) to the chosen major which could well explain differences between majors in terms of difficulty. Motivation, while not studied here, may be the only relevant difference.

One reason for the inadequacy of these probabilities involves the criterion employed to test adequacy. In either regression or discriminant analysis the customary criterion employed in tests of significance deals with the null hypothesis that random factors just by chance alone could explain the apparent relationships drawn out by certain optimizing procedures. While tests of significance are associated with the degree of relationship, they do not answer directly the question involving the usefulness of an apparent relationship. For example, one might have a very large sample giving a multiple correlation of 0.2 which would be judged significantly different from zero. This would provide, however, very little confidence in a prediction system. The criterion employed here dealt with the adequacy of a probability computed for each individual when compared to the true probability which takes on only the values of zero or one. If the average deviation between the computed probability and the true probability could be reduced as much as 0.1 by the use of certain measures, then the computed probability based upon these measures would seem to be useful. That is, improvements affecting the first decimal of the probability might be considered worthwhile. In the section of this study dealing with the probabilities of choosing the various majors the use of the measures barely improved upon the second decimal of the *a priori* probabilities. In the section dealing with the probability of graduation, the use of the measures did achieve an improvement over the *a priori* probability that affected the first decimal. The effort to compute the probability of graduation conditional upon the choice of major showed improvement over the unconditional probability that barely affected the second decimal.

The import of these findings is not that differences do not exist between these various majors on the measures studied, rather it is that efforts to employ these differences to compute probabilities for use in the decision process show such use to be questionable. The "noise" inherent in such computations is almost of the same magnitude as the differences used. While this sounds like a study that failed, it actually has considerable significance. It means that observed differences between college majors on the measures studied have less relevance to the choice of a student than would be expected. Were we to use a complete decision-making model as developed by statistical decision theorists, the measures would have little effect on the decision.

The full import of this study takes several directions. There are often inclinations to attach significance to the interaction between different measures in the context of assessing an individual's record. We might, for example, be inclined to tell a student with lower grades in chemistry than in his other subjects that he should stay away from a major involving much chemistry in the curriculum. Empirically, however, there is no evidence that students have done this uniformly in the past. We might also be inclined to tell the student that his chances of success in a chemistry-oriented major would be poorer than his chances in other majors, but this cannot be validated empirically either. An apparent reason for the absence of such validations is the fact that chemistry grades measure an academic endeavor which differs in name from other endeavors but which is nevertheless an academic endeavor. The implications of poor grades in chemistry are more relevant to the general academic achievement than their differential relevance to achievement in just chemistry-oriented endeavors.

An extreme example of the general validity of some measures in contrast to their differential validity involves work in the humanities. A common hypothesis regarding the prognosis of M.I.T. student records leads to the emphasis of measures of achievement in scientific fields as more important than measures of achievement in humanities. It seems reasonable to postulate achievement in science as a necessary condition for success at M.I.T. (which it is) but to consider achievement in humanities as desirable but not necessarily necessary. The statistical facts belie this for good reason. Measures of scientific achievement are often highly correlated among themselves. Further, they are less correlated with measures of non-scientific achievement. While the non-scientific measures are often less directly related to success at M.I.T. than the scientific measures, they have importance due to relevant factors they measure uniquely. In short, one can make a better prognosis of success at M.I.T. knowing a science grade and a humanities grade than he can knowing two science grades.

The fact that this study failed to uncover information contained in the academic measures studied that would be useful in choosing a career at M.I.T. raises the question of what objective data could be employed. It is possible that the first term involves too many adjustment and transition problems to be used for prognoses of the type made in this study. Later academic records do not bear on the first choice of a career, but conceivably they could help in decisions involving later changes of career choices.

A study of non-intellective factors could be made for potential use at the time of the first career choice. Such factors were excluded from this study but could be made part of another study. There is, unfortunately, no evidence available as to how effective these might be at M.I.T. We can say that this study indicates the academic measures alone are not useful for choosing careers at M.I.T.

I. Counseling and Statistics

The counseling of students is necessarily a complex process. When a student is faced with a decision it is important that a counselor make available to the student as much of his experience as is relevant. Further, the counselor must be prepared to offer guidance in the weighing of interests and abilities against the objectives of the individual. While the decision must be made by the individual student, the counselor can provide information that will assist the student in the decision process. In particular, he can provide the student with appropriate statistics.

For a particular problem in which some statistics will be important, the first job is to identify the most relevant variables. When several variables are thus identified, it becomes necessary to find the best way in which they are to be used. Often this requires a combination of variables such as a weighted average. In the full complex of guidance counseling it then becomes imperative to determine how statistics will be employed at the correct level of their relevance- not more than nor less than their relevance. Quite often counselors fail to use relevant statistics, but more often than not statistics may be considered more important than they really are.

To illustrate the problem of finding the most appropriate emphasis on statistics, consider a case in which it is desired to predict for an individual his expected level of achievement in some academic venture. Data will have been collected on some of his predecessors and subjected to regression analysis. From this analysis will come a prediction equation which gives for any one profile of scores the average achievement of all those who had that profile in the normative group. Some will have achieved above this expectation and some will have achieved below it. When we use the prediction equation for a particular individual we can tell him his prediction is his expected level of achievement, but to be honest we must tell him he may actually reach any level of achievement. We can, however, say that the likelihood of his achieving various levels does vary according to his expected level. Using his expectation, the standard error of estimate for the prediction system, and the tables for the normal curve we can give him, for example, the likelihood he will make a passing level of achievement. Among a group of students whose expected levels are all passing, there will nevertheless be some who fail. To be honest with each it is only fair to make this clear. In fact, giving each student his expected level and the chances he will not fail seems most appropriate. To begin with, his actual achievement will be determined by much more than the items used for prediction, and he should realize this. Secondly, those factors beyond those used for prediction are unknown and can only be considered by him as random. Accordingly, the use of probabilities to evaluate risks is appropriate. The probabilities are actually estimates of the proportion in the normative group with similar profiles who did not fail. Hopefully, the counselee can understand such statistics to be used in his decision. Indeed, the techniques now under study by students of statistical decision theory utilize risks and probabilities. While these techniques are presently somewhat removed from guidance procedures, they may soon become applied.* In their own right, however, probabilities seem appropriate for counseling and even necessary to insure against over-emphasis upon statistics.

Not only are probabilities more realistic in appraising different possible outcomes of a decision, but they make easier the combination of statistical or objective data with values and interests or subjective data. For example, an individual may have to choose between two educational programs and in one his chances for succeeding are higher than in the other.

If, in addition, he prefers the harder program, he must weigh his preferences against the risks. Knowing the chances he would be taking in each program in the form of the probability of not failing, he should be in a better position to make the decision which requires consideration of objective and non-objective information together. Hence, for this example, probabilities seem appropriate for evaluating risks.

Statistical information is often employed by counselors for purposes other than predicting achievement. A very common procedure is analysis of profiles. Often counselors make profile charts for individual students so as to compare an individual against typical profiles for various groups of individuals. Profile analysis endeavors to match individuals with groups and the result of the matching is given the individual as another piece of information relevant to his decision. The simplest form of profile analysis involves a set of vertical bars on which an individual's "profile" is drawn by connecting consecutive points on the bars. These are compared to group "profiles" which are available in the same format. While such profiles do contain information, it is difficult for an individual to weigh the information in conjunction with other information. Here again we may employ probabilities. If we were able to tell an individual what proportion of those with profiles exactly like his enter the first group, the second group, etc., he would have a summarization of his profile that is meaningful and realistic. In the first place, so-called typical profiles fail to describe the variations within the group jointly on all measures represented. Hence, a profile chart provides no simple indicator of how typical or atypical of a group an individual is. Measures have been proposed for signifying typicality (see Tiedeman and Sternberg,(11) which use proportions. The *centour* score associated with an individual profile is defined as the proportion of people in a group with profiles *less* typical of the group than the individual's profile. Such a score is one which the counselee can understand, and accordingly use of centours appears far superior to the use of the profile graphs as the summarization of information. In some instances it may be necessary to combine a profile analysis with another analysis and the centour may be less appropriate than another statistic. To illustrate this and many other points, let us consider an example.

---

* See Cronbach and Gleser(1) Psychological Tests and Personnel Decisions, for new approaches to personnel work based upon statistical decision theory.

Our example will be somewhat simplified but nevertheless will be realistic. Consider the case of students who must choose a college. For our example we will limit the choice to two colleges: College A and College B. (The typical student considers a handful of colleges, but comparisons can still be made by looking at pairs). Further, let us say that our statistical information involves only one measure, the Intelligence Quotient or IQ. Let us say we know the following distributions of IQ scores among the students in the two colleges.

TABLE I

HYPOTHETICAL IQ DISTRIBUTIONS FOR TWO COLLEGES

	Frequencies	
IQ interval	College A	College B
less than 110	65	2
110 to 119	65	13
120 to 129	45	30
130 to 139	20	70
140 and up	5	35
Total	200	150

Knowing an individual's IQ score we can appropriately tell him how typical he is of students in either college. A common technique employs percentiles. We might tell a student with an IQ of 140, that his score would place him in the 98th percentile in college A but only in the 77th percentile at college B. If high percentiles are his only objective, College A would seem most appropriate. However, of the 40 with IQ scores like his (140 or more) only 5 chose College A while 35 chose College B. It would seem appropriate to point this out to the student. He might have concern, however, for the two different standings and the apparent difference in the risks at the two colleges. For this reason, we might like a breakdown according to IQ of those graduated and those failing to graduate at each college. The following might be such a breakdown.

TABLE II

HYPOTHETICAL IQ DISTRIBUTIONS FOR TWO COLLEGES
BY GRADUATION AND FAILURE

	College A			College B		
IQ interval	grad.	fail	total	grad.	fail	total
less than 110	20	45	65	1	1	2
110 to 119	30	35	65	4	9	13
120 to 129	30	15	45	15	15	30
130 to 139	15	5	20	50	20	70
140 and up	5	0	5	30	5	35
	100	100	200	100	50	150

We can now estimate the risks for the student having an IQ of 140 at either of the two colleges. At College A 5 out of 5, or 100%, of those with IQ's of 140 or more graduated. At College B 30 out of 35, or 86%, of those at that IQ level achieved graduation. Just as did the percentile statistic, this statistic implies College A seems best in terms of risk. But, as before, this ignores the fact that 35 of the 40 who chose either college and who had IQ's of 140 or more actually chose College B. In other words, something about College B appears to make it worth the extra risk as indicated by these records. If it were desired to find a single statistic that combines these conflicting bits of information and which would be meaningful to the counselee we might consider another probability or proportion. The joint condition of graduation and attendance at a college shows 5 with IQ's over 140 who attended College A and graduated and 30 who attended College B and graduated. Since there were a total of 40 with IQ's over 140 attending either college, we can say 12.5% of those having IQ's of 140 or more and entering either A or B graduated from A while 75% of the same group graduated from B. These figures, 12.5% and 75%, are a compromise between the risk figures which favor A, and the past preferences which favor B. While the compromise does favor B, it does not do so exclusively. Further, it would seem to be a statistic which is comprehensible to the counselee. As a joint probability, it can be expressed as the product of a conditional probability and a marginal probability. Each of these probabilities are also meaningful individually and potentially useful in counseling.

An opposite example would involve a student with an IQ below 110. At College B 50% (1 out of 2) with such IQ's graduate, but small numbers make this an unreliable statistic, particularly when 31% with such IQ's graduated at College A. The joint probabilities at each college are 30% (20 in 67) attended A and graduated as compared to 1% (1 in 67) attended B and graduated. This case further illustrates how the joint probability figure tends to involve larger denominators than the risk figures. Accordingly, the use of the joint statistic may mean more reliability in the use of proportions from past experience to estimate probabilities.

On the premise that statistics of the type discussed here are worthy of study, it becomes important to know how to generalize their determination for cases more complicated than the simple example above. In the first place we might improve on the counting procedures used in the example by employing algebraic expressions of density functions. If we knew the IQ scores at each college were normally distributed with known means and standard deviations, we can employ shorter intervals than those given above. For example, instead of counting all those with IQ's of 140 and more to find how many "like" a student with an IQ of 140 choose either college, we can use density functions to estimate how many at either college had IQ's between 140 and 140 + Δ, where Δ is a very small interval of IQ scores. Denoting the mean IQ at College A by M_A and the standard deviation of IQ scores at College A by $S.D._A$, we would use the expression for the normal curve to obtain the proportion of students at College A with IQ's between 140 and 140 +Δ as:

$$(1.1) \qquad \frac{1}{\sqrt{2\pi}} \frac{1}{(S.D._A)} e^{-\frac{1}{2}\left(\frac{140-M_A}{S.D._A}\right)^2} . \Delta \quad *$$

* (see any introductory textbook in mathematical statistics for explanations of such curves. One such text is Mood's Introduction to the Theory of Statistics [6]).

With more generality we might let x denote any IQ score and dx (the differential) the small interval, and use the following to estimate the proportion at College A with IQ's between x & x + dx

$$(1.2) \qquad \frac{1}{\sqrt{2\pi}} \quad \frac{1}{(S.D._A)} \quad e^{-\frac{1}{2}\left(\frac{x-M_A}{S.D._A}\right)^2} \quad dx$$

Since a total of 200 entered College A we can estimate the number (possibly a fraction rather than a whole number) of those at College A with IQ's between x and x + dx by multiplying (1.2) by 200.

By analogy, we can express the proportion of those at

College B with IQ's between x and x + dx as

$$(1.3) \quad \frac{1}{\sqrt{2\pi}} \frac{1}{(S.D._B)} e^{-\frac{1}{2}\left(\frac{x-M_B}{S.D._B}\right)^2} dx$$

Where M_B and $S.D._B$ are the mean and standard deviation of IQ scores at College B. Multiplying (1.3) by 150 gives an estimate of the number at College B with IQ's between x and x + dx.

Consequently, the first improvement we can make on the counting procedure is to use the density functions to estimate how many with IQ's of x went to College A and how many went to B. Just as we did in the counting procedure, we can add these for the denominator of the calculated proportions. Doing so, we obtain as the proportion of those attending college A (when A or B is chosen) having IQ's of x as

$$(1.4) \quad \frac{200 \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{(S.D._A)} e^{-\frac{1}{2}\left(\frac{x-M_A}{S.D._A}\right)^2} dx}{200 \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{(S.D._A)} e^{-\frac{1}{2}\left(\frac{x-M_A}{S.D._A}\right)^2} dx + 150 \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{(S.D._B)} e^{-\frac{1}{2}\left(\frac{x-M_B}{S.D._B}\right)^2} dx}$$

in which the $\sqrt{2\pi}$ and the dx terms cancel out. This expression is a specific case of the use of Bayes' Rule. Since, in the past, Bayes' Rule has been discredited (see Fisher, 2) the foregoing exposition was given to show no tricks are involved in its use in this context. While it is true the term probability has been loosely identified with proportions in a context where rigorous probabilities are truly zero or one, proportions are themselves valid here. Calling them probabilities occasionally is a breach of rigor but extremely practical for the implications of the term.

It is desired to study the generalization of (1.4) to more than one variable and to more than one group, for that would enable us to compute one part of the desired joint probability with complete flexibility. The probability densities can be generalized readily, and the denominator can be composed of more than the two terms corresponding to the two groups. Tatsuoka(10) demonstrates the generalizations and shows how they may be combined with probabilities of success to obtain the desired joint probability.

For the probabilities derived by (1.4) or some generalization to be meaningful the groups must be exhaustive. For instance, in the context of choosing a college the groups must exhaust all who have attended college as well as those who did not. Even by grouping colleges, this requires a substantial number of groups because the situation has so many alternatives. It is the purpose of this study to investigate the feasibility of developing the joint probability statistic in a context which is less open. Specifically, it is desired to use college majors to define the groups (rather than the college attended). Graduation, as in the example, will serve as the criterion of success. The groups studied are students in the classes of 1956, 1957, and 1958 at the Massachusetts Institute of Technology. Since majors tentatively are selected by M.I.T. students near the end of the first year, the data most available at that time include first semester grades and information from admissions applications. The objective, therefore, is to use these variables to estimate optimally the probability a person will select a given major and the probability, given his selection of that major, that he will be graduated in normal sequence. The product of these two gives the joint probability of choosing the major and graduation. Tatsuoka(10) has developed the techniques for finding such probabilities, and the question here is whether these techniques can be optimized so as to improve upon his results. The question of measuring how effective would be the use of such a statistic would require, after its full development and validation, actual trials with a test group of freshmen with comparisons of follow-up studies with a control group. It must be remembered, however, that one purpose of such a statistic is to insure against over emphasis on objective data. Hence, an objective follow-up is rather inadequate. In this study, however, we limit ourselves to the problem of optimization. By a careful study of the classes of 1956 and 1957 we will develop the desired statistic and test its accuracy on the students in the class of 1958.

II. The Tatsuoka Procedures

The probability of membership in a group is developed from estimates of proportions in a group based upon past experience as demonstrated in Chapter I. To generalize expression (1.4) to more than two groups we need only to add corresponding terms to the denominator until we have all the estimates of how many in each group have scores between x and x + dx. We might denote the general denominator as follows.

$$N_A\, f\,(x|A) + N_B\, f\,(x|B) + \ldots + N_G\, f\,(x|G) \tag{2.1}$$

where $f\,(x|A)$ denotes the probability density function for x scores in group A. If we want the proportion of the total who are in group G, say, we use

$$\frac{N_G\, f\,(x|G)}{N_A\, f\,(x|A) + N_B\, f\,(x|B) + \ldots + N_G\, f\,(x|G)} \tag{2.2}$$

At this point it is clear that the density functions are not limited to the univariate case; they could just as well be multivariate. In most situations where terms of this type are to be used the densities will be multivariate since too little information is contained in only one variable. Such is the case in this study where several variables will be employed.

Under the condition that the dispersion matrices of the various group densities are all equal, the techniques of discriminant analysis are appropriate to reduce the dimensionality of the space without sacrificing information. For instance, the densities in(2.2)may involve 12 variables - a 12 dimensional space - but investigations may show a two dimensional sub-space to contain all the relevant discriminating information. Consequently, it may be possible to employ in practice only two weighted averages of the original twelve variables. Tatsuoka(10) showed that in such a case the densities in (2.2) could involve only the two weighted averages and give the same value for the proportion. The saving in computation when many proportions are to be calculated for different values of the original variables can be considerable.

Rao (7) shows how to determine the appropriate weighted averages in a discrimination problem. The weights desired are solutions to the following matrix equation.

(2.3) $(A-\lambda W)\ V = 0$

where A is the matrix of dispersions among the group centroids weighted by their size (number in the group), W is the common matrix of dispersions within groups, λ is an unknown scalar which is the latent root associated with the latent vector, V, of desired weights. Without loss of generality A and W can be matrices of sums of squares and cross-products, and we will deal with them as such.

There will be several vectors of weights, V, that will satisfy (2.3) but in discriminant analysis they cannot exceed the number of original variables or one less than the number of groups, whichever is smaller. Generally, this limit is one less than the number of groups, since it is unrealistic to use less variables than groups. Some of these solutions may be of no consequence, however, and we may have many less than the limit which will include essentially all the relevant information. If this be the case, we can attain greater efficiency in calculating proportions like (2.2).

We also desire another probability which will go with the probability of membership in a group to give the desired joint probability statistic discussed in Chapter I. This is the conditional probability of success given membership in the group. This condition is by no means a restrictive one because its development will come from an analysis of past members of the group. Indeed, in cases where chances of success can differ markedly from one group to another we would choose to employ the condition. Tatsuoka(10) used the methods of regression analysis to develop his conditional probability by means of the familiar expectancy table. To estimate the probability of graduation he determined a regression equation using the four year grade average as the criterion. Then using the standard error of estimate he computed for any predicted four year average the probability the actual four year average may exceed the minimum required for graduation. Two major disadvantages hinder this approach. The first is that the regression analysis can be made only on those who complete the fourth year. This means all those failing to reach the fourth year cannot be studied and it is only hoped an extrapolation will adequately cope with them. Further, attrition occurs at various times in the basic four year program. While attrition surely is correlated with the "hypothetical" four year average should it be achieved, disqualification may occur because one semester was worthless while others were marginal but satisfactory. In other words, the hypothetical dimension that has graduates all on one side of a value and non-graduates on the other may exist, but the four year average is not precisely that scale.

Because of the disadvantages given above it was decided to apply discriminant techniques to this problem as well by considering the prediction of membership in either of two groups: those graduated and those not graduated. The procedure of Tatsuoka's would be completed as well and a comparison made empirically.

But this decision, while providing a better logical basis for correcting certain faults, created others due to the fact that the basic assumptions usually made are known not to be true. Not only is it not true that the density functions such as those given in (2.2) have identical dispersion matrices, but the assumption that they are normal density functions is not even a good one. Accordingly, several different procedures were attempted in some pilot studies to see if some means could be found that would be better than the conventional. These are discussed in Chapter III.

III Pilot Studies: Probability of Success

The use of a linear form of t variables in regression analysis means that all points in the t-space which give the same value of the linear form lie in a hyper-plane of t-1 dimensions. In the case of two independent variables this means the points in the two-dimensional space which all give the same value for the linear form lie on a straight line orthogonal to the linear form. Since the usual linear form is a weighted average of two predictors this means a low score on one predictor may be counter-balanced by a high score on the other. In theory, this fact is disturbing since a low score in one variable may be indicative of real difficulty almost regardless of another score. For instance a student may have an extremely low verbal aptitude score but have a good math aptitude score. It seems quite reasonable that his potential should be judged less than a student with average scores in both variables. In practice this consideration is minor by virtue of the high correlation between most educational measures. We hardly ever find widely divergent scores, and when we do, errors of measurement could likely have been the cause. Nevertheless, it was decided to investigate the possibility that a non-linear model might be appropriate in deriving the probability of success.

Since one set of variables that would be available in the counseling situation at M.I.T. are the grades in the first semester, some tabulations were made on two grades at a time against graduation. For each possible combination of two letter grades the number graduated and the number failing graduation were found. Imagine, then, a three-dimensional surface which has for its base the two scales of grades. The third or vertical dimension we would make the proportion who graduate having the corresponding grades read from the scales at the base. We can then find the intersection of this surface with horizontal planes at different levels of the proportion. In the case where linear forms are appropriate, these intersections will all be straight lines. If, however, the linear form is inappropriate, the intersection will be a curve. Since the intersection involves all the grade combinations having the same proportion graduate, it would be best to use a prediction function of the same type as the intersection.

An empirical approximation was made to the surface described above by determining for each combination of grades the proportion who graduate. Using straight lines to connect adjacent points, it was possible to approximate the surface. The intersections were then found for the proportions graduating of .25, .50, and .75. As would be expected, the curves of intersection were far from being smooth curves. They did, however, show evidence of being non-linear. The smoothest and best appearing of the curves is shown in Figure 1. Since this empirically determined curve connects the grade combinations having the same potential for graduation, it is desired to find more accurately from the data the constants for such a curve. Since the empirical curves have the characteristics suggested by the concern for the inadequacy of linear forms discussed earlier, it was decided to pursue the determination of such curves.

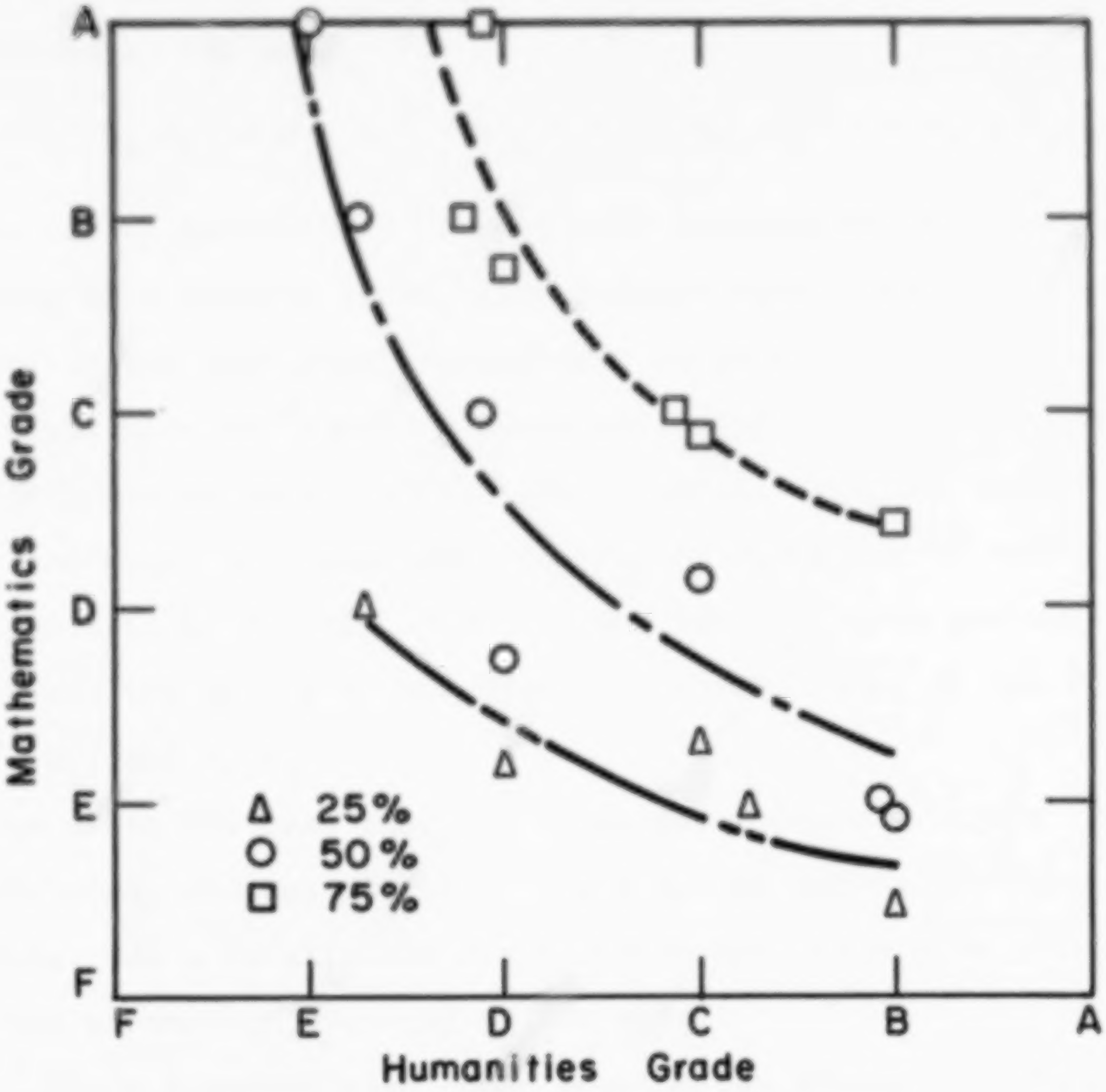

FIG. I EMPIRICAL CURVES FOR POINTS HAVING 25%; 50%; AND 75% OF STUDENTS GRADUATING

Since the characteristics of the empirical curves shown in Figure 1 are similar to conic sections, it is possible to consider the use of a generalized quadratic as the model for the curves. With two independent variables we might write the function as follows:

$$(3.1)\ f(x_1,x_2) = a_{11}\ x_1^2 + a_{12}\ x_1x_2 + a_{22}\ x_2^2 + b_1x_1 + b_2x_2 + c$$

This can be generalized to many more independent variables by adding more squared terms, more product terms, and more linear terms in the additional variables. In general for K variables we would have $(K^2+3K+2)/2$ terms including the constant and accordingly as many coefficients to solve. With 10 variables, for instance, this approach requires 66 terms and 66 coefficients to be solved. Further, the least squares process involves the matrix of correlations between terms of the form x_i^2, x_i, and $x_i\ x_j$.

It is often the case that the first two types are highly correlated, and the matrix of correlations approaches singularity. As a result, the coefficients are often poor estimates of the parameters proposed by the model.

Since a generalized conic section can always be rotated to a form involving no product terms, it appeared possible to consider models involving rotations of axes. This approach would require, then, the solution of centroids and major axes in the rotated space. As an example, the model in equation (3.1) could be rewritten as follows:

$$(3.2)\ f(x_1,x_2) = \alpha_1\ (y_1-\beta_1)^2 + \alpha_2\ (y_2-\beta_2)^2$$

where α and β are now constants which determine the major axes and the centroid of the conic section in terms of y_1 and y_2, which are the bases for the rotated space. There is no reason for requiring that the rotation be orthogonal, but it could be.

Another approach involves the use of probability density functions in the same manner described in the previous chapter using only two groups, one being those graduated and the other those failing to be graduated. With known density functions the curve of intersection discussed earlier becomes

$$(3.3) \qquad \frac{f\,(\,x \mid A)}{f\,(\,x \mid B)} = \text{constant}$$

where A and B denote the graduated and not graduated groups. In the special case where the two density functions are normal densities, (3.3) becomes

$$(3.4)\ \frac{f(x|A)}{f(x|B)} = \frac{C_A \exp[Q_A(x)]}{C_B \exp[Q_B(x)]} = \frac{C_A}{C_B} \exp[Q_A(x) - Q_B(x)] = \text{constant}$$

where $Q_A\,(\,x)$ and $Q_B\,(\,x\,)$ are the appropriate quadratic forms for each density function. When the density functions are multivariate normal, the quadratic forms can be expressed in matrix notation as

$$(3.5)\ Q_A\,(x) = (\,x - \mu_A)'\ C_A\ (\,x - \mu_A)$$

$$(3.6)\ Q_B\,(x) = (\,x - \mu_B)'\ C_B\ (\,x - \mu_B)$$

where x is the column vector of scores.

μ_A, μ_B are group mean column vectors

C_A, C_B are each the inverse matrix of the group variance - covariance matrices.

Taking the logarithm of equation (3.4) gives the equipotential surface as

$$(3.7)\ \varphi = Q_A(x)-Q_B(x) = (x-\mu_A)'\ C_A(x-\mu_A) - (x-\mu_B)'\ C_B(x-\mu_B) = \text{constant}$$

which is a hyper-conic section when C_A is not equal to C_B and a hyperplane when it is.

Despite the fact that the assumption of normality is weak when the groups are the graduated and the not graduated students, the normal density approach did approximate the empirical curves in Figure 1 as closely as the solutions to (3.1) did for the two dimensions.

As a preliminary analysis it was decided to experiment with the normal density assumption on some of the first data collected. In the first class for which data was collected (the M.I.T. Class of 1956) there were 818 students for whom each of four grades were available in the core subjects of the first semester of the freshman year. Of the 818 students, 611 were graduated, and 207 were not. The variance-covariance matrices and means for the grades in the four subjects are given in Tables III and IV.

TABLE III

Variance-Covariance: 611 Graduates, Class of 1956

	Grade			
	Chemistry	Physics	Humanities	Mathematics
Chem.	.91803	.62145	.22445	.61926
Phys.	.62145	1.06983	.21597	.63647
Hum.	.22445	.21597	.65372	.17423
Math.	.61926	.63647	.17423	1.05818
Means*	3.68903	3.31424	3.45335	3.87561

*Grades at M.I.T. are on a six-point scale on which A is 5, B is 4, C is 3, D is 2, E(conditional failure) is 1, and F(failure) is 0.

TABLE IV

Variance-Covariance: 211 Nongraduates, Class of 1956

Grade

	Chemistry	Physics	Humanities	Mathematics
Chem.	1.54392	1.01990	.33088	1.18274
Phys.	1.01990	1.98010	.32836	1.37811
Hum.	.33088	.32836	1.16869	.35690
Math.	1.18274	1.37811	.35690	2.18985
Means	2.32836	2.00000	2.84079	2.39801

Using these variance-covariance matrices and the vectors of means in equation (3.7) it is possible to find values of ϕ, the difference between two quadratic forms, for each individual. When this model is appropriate, the ϕ values should be large in the positive direction for those in group B and large in the negative direction for those in group A. The distribution of ϕ values will be bell-shaped, but not a Gaussian distribution. Nevertheless, it should provide a better separation between the two groups if the assumptions are valid. Since we know the assumptions do lack validity in the case at hand, the question is whether the *curvilinear nature of the* ϕ *function may still* improve upon the linear function known to be somewhat inadequate. Accordingly, values of ϕ were computed for each of the 818 students in the group after transforming ϕ into a canonical form for ease of computation.

Since the possibility of a non-linear model seems quite real in the context of subject grades where a subliminal grade in one subject conceivably should not be balanced out by any other grades, attempting several approaches seemed feasible.

It seemed not too unreasonable to consider the Euclidean distance between a point representing an individual's grades and the point corresponding to the highest grades possible. Such a distance function would provide surfaces for a given value having the desired mathematical characteristics. Further, it has the advantage of insuring a monotonic function as one moves from poor grades to excellent grades. More general curvilinear models do not give this insurance.

To test the plausibility of the distance function a slight change was made which allowed for scale differences in each of the four grades which were the basis for the four-dimensional space. This allowance would enable hyper-ellipses to replace hyper-spheres for the surfaces obtained for constant values of the distance function. The equation for such a distance function would be

$$(3.8) \quad d^2 = a_1(5-x_1)^2 + a_2(5-x_2)^2 + a_3(5-x_3)^2 + a_4(5-x_4)^2$$

where $x_{i_j} i = 1,2,3,4$ are the grades

$a_{i_j} i = 1,2,3,4$ are weights related to the major axes of the ellipses.

5 corresponds to the maximum grade of A

The weights, a_i, were determined empirically by use of least squares methods using as the criterion the dummy variable whose value is 0 for graduates and 1 for non-graduates. This is equivalent to the determination of the discriminant function applied to squared terms rather than the usual linear terms. While not the best criterion, this seemed reasonable for exploration purposes. The weights resulting were

i	a_i	Grade
1	.1346	Chemistry
2	.0695	Physics
3	.0818	Humanities
4	.1141	Mathematics

While empirically derived, these weights do suggest relatively more emphasis upon chemistry and relatively less upon physics, but all the weights are of the same magnitude.

As part of the same exploration it was decided to determine the discriminant function which uses the two groups as its criterion. This involves finding the weights, c_i, in the linear form

$$c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 \tag{3.9}$$

which maximizes the "student's t" statistic for the difference between the two means on the linear form. The weights resulting were

i	c_i	Grade
1	.6411	Chemistry
2	.2612	Physics
3	.3498	Humanities
4	.5232	Mathematics

These weights show the same inter-relationship as do the weights for the distance function. That is, chemistry has the largest weight, and physics has the smallest. As before, the weights are of the same magnitude.

One "control" function to which the trial functions can be compared is the weighted grade average in which each weight is proportional to the estimate of total hours spent in class, lecture, laboratory and in preparation. These hours are

Hours	Grade
11	Chemistry
11	Physics
8	Humanities
9	Mathematics

The next question is how these various functions compare with each other in terms of separating graduates and non-graduates. Since the distribution of scores for either group on each function is not the Gaussian distribution, the "student's t" statistic is not appropriate. For demonstration purposes it was decided to rank all 818 students by deciles on each of the four functions and count within each decile the number graduating. A perfect discriminating function would have none graduating at one end of the deciles and all graduating at the other end. One decile would be a mixed one with some graduating and some not whereas the others would be pure. Table V summarizes these comparisons by deciles. For each experimental function is given the sum of the absolute differences between its numbers and those of the perfect discriminator. Also are given the number graduated in the bottom 20% of each function, which represents errors if these be tagged as likely not to graduate.

Table V

Number Graduating at Various Deciles of Experimental Functions

Decile	Number in Decile	Function Normal Density Quadratic	"Distance"	Two Group Discriminant	Grade Average	Perfect Discriminator
1	82	30	10	11	9	0
2	82	50	44	45	45	0
3	82	52	48	48	55	39
4	82	64	67	68	63	82
5	81	72	68	70	68	81
6	81	60	73	66	69	81
7	82	67	76	76	79	82
8	82	65	70	71	70	82
9	82	75	75	74	71	82
10	82	76	80	82	82	82
Sum of absolute diffs. with perfect discr.		186	126	130	140	0
Number graduated in two lowest deciles		80	54	56	54	

By either crude criterion for comparing these functions the function based upon the normal density quadratic is the poorest. The distance function makes less mistakes overall as compared to the perfect discriminator, but as a means for "identifying" those likely not to graduate it shows no real improvement over the two linear functions. In fact, none of these functions succeeds at all well in this regard. When one in three of those "tagged" not likely to graduate actually do graduate, it is questionable whether such a tag should be employed. At any rate, neither of the two curvilinear functions shows any startling improvement over linear functions when tested on the same group from which they were derived. The chance that they would appear superior when tested in another group was quite remote.

Consideration was given another quadratic model which was more flexible than the others in that less assumptions were involved. This involved finding a quadratic form expressed in matrix notations as

$$(xT - B)' D (xT - B)$$

where x is the column vector of variables used

T is a transformation matrix

B is a column vector which gives the centroid of the hyper-conic section in the transformed space

D is a diagonal matrix which determines the major axes of the hyper-conic section.

A least squares solution to the maximization of the difference between the two group averages with some minor restrictions on the T and B matrices to keep them finite produced the following matrix equation for the columns of T

$$\left\{ C_A - \theta \, [C_A + C_B + (M_A - M_B)\,(M_A - M_B)\,] - \theta^2 \, C_B \right\} \; t = 0$$

where C_A and C_B are variance-covariance matrices for the groups A and B,

and M_A and M_B are column vectors of group means,

and t is the vector which corresponds to the value of θ and which is a column of T,

and θ is an unknown scalar.

This rather over-complicated solution was determined for some simple cases, one of them being the two variables of grades in first semester humanities and mathematics. Even in the simple case this model proved inferior.

To summarize the results of these explorations, it was evident that efforts to create a model more flexible than the linear model and one which would be determined empirically were fruitless. As greater flexibility was introduced into the empirical determination the models functioned less well. The simplest, namely the distance function, showed better results than others, but not sufficiently better than linear models to warrant its use. Therefore, it was decided to employ a linear model as the basis for determining the probability of graduation.

The probabilities desired are actually conditional probabilities in that for any individual we would determine the probability of graduation given the fact that he actually chooses a specific major. When this is combined with the probability that he chooses that major, we have the joint probability discussed earlier. Since the decision to employ a linear form as the basis for the conditional probability was made, the next question involves the extent to which the parameters of the model vary between majors. The coefficients in a linear model determine the orientation of a vector in the space of the variables used. The constant term in the model determines the point in that space through which the vector must pass. The most general solution to the question is to let the direction cosines of the vector and a specific point be determined for each of the majors independently. This involves, however, the determination of coefficients and the constant term for some majors having as few as three or four students. In fact there may be several cases in which the degrees of freedom are less than the number of coefficients to be determined.

One solution to the problem would be to use a common vector for all majors but alter the constant term for the various majors. This requires some evidence of collinearity of independently determined vectors. Tests of significance do exist for this purpose (See Kempthorne, (3), pp 47-49) when the criterion is a continuous variable. For our case the criterion is the dichotomy of graduated - not graduated, which can be dealt with in a similar manner provided certain assumptions regarding the normal distribution of a residual term are valid under the two conditions of the null hypothesis and the alternate hypothesis. It was decided, however, that for our purposes a simpler demonstration would suffice. The final test of the hypothesis of a common vector requires consideration of the obvious advantages against the potential loss of information. With this in mind it would be possible to look at potential differences with the numbers of students affected to see if the net assets made the simpler case undesirable.

For purposes of this exploration it was decided to employ a technique which included all students whether they were graduated or not. As will be seen later this technique is investigated for its validity in comparison with the technique originally planned. Briefly, what is considered is the two group discriminant function using the dichotomy of graduated-not graduated as the criterion. The function resulting from this analysis specifies the direction cosines of the vector discussed earlier. The function is determined by finding the matrix product of the variance-covariance matrix and a column vector of differences between means. That is, if we were to find the discriminant function best separating graduates from non-graduates in the group G of students in a particular college major, we would find the vector

$$C_G^{-1} \quad d_G$$

where C_G is the variance-covariance matrix of scores in group G

d_G is the vector of differences between means of the two groups (graduated; not graduated)

Were we to compute a separate variance-covariance matrix for each group we would again be vulnerable to large sampling fluctuations in the small groups. Unless there is strong evidence to the contrary, which there is not, we could use the same variance-covariance matrix for each group. When we do so, we are then merely rotating many different d_G vectors into the same space. This rotation is not necessarily orthogonal. If two different d_G vectors are collinear or nearly so, they would be so in the transformed space as well, provided the transformation is well-behaved, which it is. The result of this discussion is that an indication of the degree of collinearity of different vectors for each group of majors may be obtained by investigating the collinearity of the different d_G vectors. This conforms to a common-sense approach in which one could argue for comparing the differences between the two means from one group to another. While majors may differ in their difficulty, it seems reasonable to have differences between the means being homogeneously similar but differing in their magnitudes.*

Since the quality of work in the first semester will likely be the most relevant consideration we will have, the basic variables used for this exploration are the four grades from first semester core subjects. Following are the group means and differences by college majors.

*That is, we might expect the difference between the means for graduates and the means for non-graduates to be fairly constant from one major to the other even though the means may themselves differ.

Table VI

Average Grades in Freshman Subjects -- Graduates; Non Graduates by College Major

Major	Number in this Analysis	Chemistry			Physics			Humanities			Mathematics		
		Grad.	Non Grad	Diff.	Grad.	Non Grad.	Diff.	Grad.	Non Grad	Diff.	Grad.	Non Grad.	Diff.
Civil Engineering	87	3.36	2.00	1.36	2.95	2.03	0.92	3.25	2.86	0.39	3.68	2.50	1.18
Mechanical Engineering	239	3.54	2.50	1.04	3.40	2.49	0.91	3.26	2.89	0.37	3.86	2.86	1.00
Metallurgy	54	3.64	2.61	1.03	2.86	1.94	0.92	3.40	2.72	0.68	3.78	2.71	1.07
Architecture	55	3.40	2.21	1.19	3.00	1.73	1.26	3.48	2.90	0.58	3.68	2.56	1.12
Chemistry	38	3.93	3.09	0.84	3.15	1.09	2.06	3.44	2.82	0.62	3.74	2.64	1.10
Electrical Engineering	239	3.58	2.46	1.12	3.43	2.42	1.01	3.30	2.59	0.71	3.91	3.00	0.91
Biology	39	4.04	3.00	1.04	3.58	1.83	1.75	3.81	3.17	0.64	4.00	2.67	1.33
Physics	130	4.40	2.86	1.54	4.28	3.05	1.23	3.69	3.00	0.69	4.58	3.53	1.05
General Science	12	4.43	2.33	2.10	3.50	2.25	1.25	3.50	3.00	0.50	4.75	2.75	2.00
General Engineering	18	3.54	2.80	0.74	3.31	2.00	1.31	3.15	2.60	0.55	4.00	2.40	1.60
Science Teaching	5	4.00	----	----	3.00	----	----	3.60	----	----	4.00	----	----
Chemical Engineering	203	4.06	2.63	1.43	3.30	1.80	1.50	3.45	2.51	0.94	3.96	2.26	1.70
Geology	42	3.55	2.55	1.00	3.35	2.36	0.99	2.89	2.95	-0.06	3.70	3.00	0.70
Naval Architecture	20	3.36	1.78	1.58	3.00	2.22	0.78	3.00	2.22	0.78	3.91	2.00	1.91
Economics	27	3.67	2.67	1.00	3.45	1.83	1.62	3.52	3.67	-0.15	3.81	3.50	0.31
Industrial Management													
Physical Sciences	134	3.35	2.31	1.04	3.11	2.11	1.00	3.53	2.98	0.55	3.76	2.77	0.99
Chemical Sciences	43	3.65	3.13	0.52	2.85	1.95	0.90	3.50	3.17	0.33	3.75	2.77	0.98
Aeronautical Engr.	90	3.39	2.13	1.26	3.20	1.97	1.23	3.31	2.74	0.57	3.80	2.76	1.04
Building Construction	21	3.44	1.80	1.64	3.12	1.00	2.12	3.06	3.00	0.06	3.56	2.20	1.36
Mathematics	49	4.52	3.38	1.14	4.32	2.83	1.49	4.12	3.21	0.91	4.70	4.12	0.58
Meteorology	2	3.50	----	----	1.50	----	----	4.00	----	----	3.50	----	----
Food Technology	13	3.67	3.50	0.17	2.89	1.00	1.89	3.44	2.75	0.69	3.67	3.00	0.67
Bio Chemical Engineering	5	3.40	----	----	3.00	----	----	3.20	----	----	3.40	----	----

As one means for comparing the vectors we can find the cosine of the angles between any two in the space of the four grade scores. While the cosines will likely change under the transformation, they would not be altered drastically. Such cosines are given below for some of the possible pairs.

Table VII

Cosines Between Some of the Difference Vectors From Table VI

	Civil Engr.	Mech. Engr.	Metallurgy	Arch.	Chem.
Mechanical Engr.	.99	---	---	---	---
Metallurgy	.98	.99	---	---	---
Architecture	.98	1.00	.99	---	---
Chemistry	.87	.91	.91	.95	---
Electrical Engr.	.97	.98	.99	.99	.92
Biology	.93	.97	.96	.99	.99
Physics	.99	.99	.98	.99	.91
General Science	.99	.99	.97	.97	.84
General Engr.	.93	.95	.96	.96	.94
Chem. Engr.	.97	.99	1.00	.99	.93
Geology	.95	.95	.90	.95	.90
Naval Arch.	.97	.96	.96	.94	.80
Economics	.80	.82	.76	.85	.90
Industrial Management					
Phys. Sciences	.98	.99	.99	1.00	.93
Chem. Sciences	.93	.96	.96	.97	.95
Aeronautical Engr.	.99	1.00	.99	1.00	.94
Bldg. Construction	.94	.95	.91	.96	.95
Mathematics	.89	.92	.93	.96	.95
Food Technology	.69	.75	.77	.82	.96

The cosines in Table VII are in most cases close to unity, which means collinearity. In Table VII the lowest cosines involve the students majoring in economics and in food technology.

The two groups together number 40 students. Such divergences as these were deemed minor in light of the overwhelming evidence for collinearity. Hence, the conclusion was to use a common vector for all the college major groups but to include differences between majors in the constant term of the complete model.

A question raised but unanswered is whether the use of discriminant analysis is appropriate for determining the linear form to be used. The more customary procedure (and the one planned initially) would use the methods of regression analysis to predict the average of the full four years. This implies that the four year averages do in fact serve as the best single known variable that distinguishes between those graduating and those not graduating. In reality there is no single variable that meets these specifications, since success in graduating involves several sequential steps, namely satisfactory studies over at least eight semesters. While a cumulative average may aid in the examination of an individual student's record it does not serve as the basis for the sequence of actions. Perhaps the worst indictment of the regression procedures is the fact that only those who complete the full four years are included in the analysis. Those failing to reach the end of the fourth year may have some distinctive characteristics which obviously would miss detection. The hope is that an analysis of four year records would enable one to extrapolate "backwards", but that seems to be only a hope. For the above reasons the use of discriminant procedures seem more desirable. One shortcoming with such procedures is the fact that the criterion for their derivation (student's "t" test) tacitly assumes the variances of the two groups are equal. This, of course, we know not to be the case.

It was felt that the best way to decide which approach to use would be to try both and merely see which is best. Consequently, we are making two independent determinations of the basis for the conditional probabilities. The job of choosing between them can be done empirically.

IV. Clustering of Career Majors.

There were nineteen departments of the Institute offering undergraduate programs of study at the time the classes of 1956 and 1957 were in attendance. In some of these there were options with different curricula. Altogether it would be possible to isolate about thirty curricula bearing different names. Often, however, it is difficult to distinguish between the programs of students in two different "careers". Even more often it is difficult to distinguish between the students in two different careers except by a registration card. Many overlaps exist in terms of academic measures as well as in indiviual interests or goals.

If we were to try to estimate parameters for use in discriminating between some thirty groups using the records of some 1200 students, the maximum size of the smallest group we could hope to find would be forty. Even if this were likely (and it is certainly not) we would be estimating parameters with small efficiency. That is, there would be large standard errors for the coefficients. In addition, it is not reasonable to use discriminant techniques with less variables than groups.

The foregoing suggests combining some career groups for reasons related to the analysis. Since many careers are so similar already it is realistic to consider combining some for reasons related to the groups themselves. The techniques employed for making these clusterings of departments are described in detail in Rao (7) Chapter 9. Briefly, the procedures find geometric distances between pairs of group centroids in which the distances are corrected for correlations between all variables and the scales are all made dimensionless. These are the distances, D^2, of Mahalanobis. By comparing all the pairwise distances it is possible to identify those groups which are close to each other in the geometric sense but far from other groups. These would constitute a cluster of groups which are relatively indistinguishable in the space of the variables involved. In our problem a cluster would combine career groups which are indistinguishable in terms of academic measures. While other variables might make distinctions between groups within a cluster, it is sufficient for our purposes to allow them to be grouped by academic variables. If a counselee is told there are no differences of an academic nature he should realize his private distinctions or preferences should be dominant in his decision.

In a practical sense the clustering of departments by similarity on academic measures will mean we use the same probability value for each career group in a cluster. As described above, this can and should be realistic. If we find differences between departments in a cluster in terms of the difficulty in achieving graduation, we can use different values for the conditional probability of graduation. Hence, we will use the clustering procedure only for the discriminant analysis.

Before computing all the various D^2 values possible from some thirty groups, some of the options were clustered by different criteria. In three of the engineering departments (electrical, mechanical and aeronautical) there were special co-operative programs which thereby bear a different name. The curricula in these programs differ largely because of the work experiences. Still they can be combined for our purposes, particularly since these programs do not start until the junior year.

As the basic set of dimensions for the geometric analysis we chose academic measures. Since there are no restrictions upon how many or how few of these we can employ, it was decided to use eight grades from the first year, two each in chemistry, physics, humanities, and mathematics. Half of these are ones to be included in our discriminant analysis and half are not to be included. In just the same manner as illustrated by Rao (7) in an example (pp. 357-364) we computed the different distances in terms of the eight grades and for each group ranked the other groups by increasing distance. For simplicity we have employed numerical designations for each group that are employed by M.I.T. for internal purposes. The students studied here were in the combined classes of 1956 and 1957.

Table VIII

Values of D^2 (Based on 8 M.I.T. Grades) Arranged in Increasing Order of Magnitude

1		2		3		4A		5		6		7		8	
Civil Engineering		Mechanical Engineering		Metallurgy		Architecture		Chemistry		Electrical Engineering		Biology		Physics	
N = 80		N = 237		N = 53		N = 52		N = 37		N = 328		N = 40		N = 131	
15A	0.229	9B	0.161	10	0.203	16	0.189	7	0.849	2	0.220	9C	0.514	18	0.354
4A	0.282	15A	0.209	1	0.311	15A	0.213	10	0.871	14	0.288	10	0.607	6	0.405
3	0.311	6	0.220	9B	0.384	17	0.223	3	0.967	9B	0.306	20	0.686	14	0.554
16	0.323	16	0.326	4A	0.416	1	0.282	9C	1.128	15A	0.341	14	0.717	9B	0.899
9B	0.338	17	0.342	6	0.461	3	0.416	20	1.311	1	0.365	15B	0.809	10	0.984
6	0.365	1	0.397	15B	0.475	15B	0.454	14	1.543	8	0.405	5	0.849	2	1.039
2	0.397	13	0.538	15A	0.488	2	0.595	15B	1.613	16	0.411	6	0.917	3	1.127
14	0.516	3	1.539	2	0.539	10	0.678	6	1.690	10	0.450	3	0.958	9C	1.176
17	0.573	20B	0.577	14	0.578	6	0.682	8	1.775	3	0.461	18	1.061	7	1.187
13	0.628	4A	0.595	17	0.642	9B	0.696	9A	1.855	18	0.545	4A	1.126	1	1.187
10	0.637	10	0.608	16	0.672	13	0.763	1	1.856	17	0.580	8	1.187	15A	1.258
20B	0.824	14	0.653	9A	0.825	14	0.820	4A	1.944	4A	0.682	17	1.269	9A	1.265
15B	0.839	12	0.795	7	0.958	7	1.126	18	2.019	7	0.917	15A	1.294	16	1.509
9A	1.049	15B	0.860	5	0.967	9A	1.225	9B	2.217	15B	0.972	16	1.324	20B	1.765
18	1.113	8	1.039	20B	0.994	20B	1.301	15A	2.393	13	1.002	9A	1,384	5	1.775
8	1.187	18	1.070	18	1.042	12	1.374	2	2.476	9A	1.024	1	1.384	17	1.803
7	1.384	9A	1.136	8	1.127	18	1.662	17	2.494	20B	1.050	2	1.577	4A	1.860
12	1.386	7	1.577	13	1.171	20	1.813	16	2.545	9C	1.456	9B	1.693	15B	2.079
9C	1.853	5	2.476	12	1.410	8	1.860	20B	3.449	12	1.546	20B	2.751	13	2.233
5	1.856	9C	2.556	9C	1.587	9C	1.903	13	3.813	5	1.690	13	2.852	12	2.619
20	2.513	20	2.696	20	1.725	5	1.944	12	4.276	20	2.174	12	3.576	20	3.102
19	6.866	19	8.004	19	5.966	19	5.549	19	4.851	19	7.022	19	3.601	19	8.116

Table VIII (cont'd.)

9A		9B		9C		10		12		13		14		15A	
General Science		General Engineering		Science Teaching		Chemical Engineering		Geology, Geophysics		Naval Architecture		Economics		Industrial Mgt. (Phys. Sciences)	
N = 13		N - 20		N = 7		N = 198		N = 41		N = 21		N - 29		N = 130	
10	0.533	2	0.161	7	0.514	3	0.203	13	0.624	9B	0.484	6	0.288	16	0.128
9B	0.671	6	0.306	5	1.128	15B	0.400	9B	0.687	16	0.526	18	0.326	2	0.209
3	0.825	20B	0.321	8	1.176	6	0.450	20B	0.782	2	0.538	15A	0.464	4A	0.213
20B	0.860	1	0.338	18	1.262	9B	0.506	2	0.795	17	0.608	1	0.516	1	0.229
6	1.024	15A	0.369	14	1.276	9A	0.533	15A	1.002	15A	0.623	8	0.554	17	0.242
1	1.049	3	0.384	10	1.313	7	0.607	17	0.065	12	0.624	3	0.578	6	0.341
2	1.136	13	0.484	6	1.456	2	0.608	16	1.247	1	0.628	2	0.653	9B	0.369
15A	1.206	10	0.506	9A	1.538	1	0.637	4A	1.374	20B	0.719	10	0.693	14	0.464
15B	1.206	16	0.558	3	1.587	17	0.676	1	1.386	4A	0.763	7	0.717	3	0.488
4A	1.225	17	0.589	20	1.742	4A	0.678	3	1.410	6	1.002	9B	0.762	13	0.523
8	1.265	9A	0.671	1	1.853	14	0.693	6	1.546	3	1.171	16	0.766	15B	0.669
17	1.314	12	0.687	4A	1.903	15A	0.739	9A	1.689	9A	1.350	4A	0.820	10	0.739
13	1.350	4A	0.696	15B	2.015	16	0.857	10	1.775	10	1.402	15B	0.934	20B	0.825
14	1.361	14	0.762	16	2.094	5	0.871	15B	2.005	15B	1.608	17	1.018	12	1.002
7	1.384	8	0.899	15A	2.120	8	0.984	14	2.316	14	1.824	9C	1.276	18	1.070
18	1.436	18	0.965	9B	2.260	18	0.991	8	2.619	8	2.233	9A	1.361	9A	1.206
16	1.462	15B	0.989	17	2.322	20B	1.112	18	2.823	18	2.384	20B	1.398	8	1.258
9C	1.538	7	1.693	2	2.556	20	1.254	7	3.576	7	2.852	5	1.543	7	1.294
12	1.689	5	2.217	13	3.480	9C	1.313	5	4.276	9C	3.480	13	1.824	9C	2.120
5	1.855	9C	2.260	20B	3.519	13	1.402	9C	4.570	5	3.813	20	2.146	5	2.393
20	2.789	20	2.986	12	4.570	12	1.775	20	4.965	20	3.980	12	2.316	20	2.472
19	7.977	19	8.457	19	4.634	19	5.802	19	11.332	19	9.905	19	5.657	19	6.666

Table VIII (cont'd.)

15B		16		17		18		19		20		20B	
Industrial Mgt. (Chem.Sci.)		Aeronautical Engineering		Building Construction		Mathematics		Meteorology		Food Technology		Biochemical Engineering	
N = 41		N = 86		N = 22		N = 52		N = 2		N = 14		N = 5	
10	0.400	17	0.124	16	0.124	14	0.326	20	2.355	7	0.686	9B	0.321
4A	0.454	15A	0.128	4A	0.223	8	0.354	7	3.601	15B	0.992	2	0.577
3	0.475	4A	0.189	15A	0.242	6	0.545	15B	4.259	10	1.254	13	0.719
17	0.478	1	0.323	2	0.342	9B	0.965	9C	4.634	5	1.311	12	0.782
15A	0.669	2	0.326	15B	0.478	10	0.991	5	4.851	3	1.725	1	0.824
16	0.775	6	0.411	1	0.573	3	1.042	4A	5.549	9C	1.742	15A	0.825
7	0.809	13	0.526	6	0.580	7	1.061	14	5.657	17	1.813	9A	0.860
1	0.839	9B	0.558	9B	0.589	15A	1.070	10	5.802	4A	1.813	3	0.994
2	0.860	3	0.672	13	0.608	2	1.070	3	5.966	14	2.146	6	1.050
14	0.934	14	0.766	3	0.642	1	1.113	17	6.223	6	2.174	10	1.112
6	0.972	15B	0.775	10	0.676	9C	1.262	18	6.364	16	2.183	17	1.195
9B	0.989	10	0.857	14	1.018	16	1.413	16	6.482	19	2.355	16	1.249
20	0.992	12	1.247	12	1.065	9A	1.436	15A	6.666	15A	2.472	4A	1.301
9A	1.206	20B	1.249	20B	1.195	15B	1.502	1	6.866	1	2.513	14	1.398
20B	1.400	7	1.324	7	1.269	20B	1.549	6	7.022	18	2.682	15B	1.400
18	1.502	18	1.413	9A	1.314	17	1.643	9A	7.977	2	2.696	18	1.549
13	1.608	9A	1.462	18	1.643	4A	1.662	2	8.004	9A	2.789	8	1.765
5	1.613	8	1.509	8	1.803	5	2.019	8	8.116	9B	2.986	7	2.751
12	2.005	9C	2.094	20	1.813	13	2.384	9B	8.457	8	3.102	5	3.449
9C	2.015	20	2.183	9C	2.322	20	2.682	20B	9.905	13	3.980	9C	3.519
8	2.079	5	2.545	5	2.494	12	2.823	13	9.905	20B	4.271	20	4.271
19	4.259	19	6.482	19	6.223	19	6.364	12	11.332	12	4.965	19	9.905

One observation of the many D^2 values in Table VIII is the fact that the various departments differ very little in the measures which serve as the basis of the eight dimensional space. Nevertheless, it seemed advisable to make the clusterings which would combine those majors most homogeneous in the eight dimensions. The following clusters were made from study of the distances.

Group A (Engineering based upon physics)

1 Civil Engineering
2 Mechanical Engineering
9B General Engineering
12 Geology and Geophysics
13 Naval Architecture
15A Industrial Management based upon Physical Sciences
16 Aeronautical Engineering
17 Building Construction
20B Biochemical Engineering

Group B (Engineering based upon chemistry)

3 Metallurgy
9A General Science
10 Chemical Engineering
15B Industrial Management based upon Chemical Sciences

Group C

6 Electrical Engineering

Group D

8 Physics
14 Economics and Science or Engineering
18 Mathematics
19 Meteorology

Group E

5 Chemistry
7 Biology
9C Science Teaching
20 Food Technology

Group F

4A Architecture

Meteorology, with its two students, could not really be classified in any cluster accurately since two people are too few for any estimation. In the meantime this undergraduate program has been discontinued, so the practical considerations make the assignment academic.

Many other assignments seem rather arbitrary, and indeed they are. It can be said, however, that the objective material did serve as the basis for the clusterings. In Table IX are given the inter-and intra-cluster mean distances weighted by the product of the numbers in each department of the two involved in a distance computation.

Table IX

Inter-and Intra-Cluster Average Distances, D^2

	A Eng:Phys.	B Eng:Chem.	C Eng:Elec.	D Sc:Phys.	E Sc:Chem.	F Arch.
A	.417	.780	.420	1.307	2.197	.475
B	.780	.366	.547	1.152	.986	.626
C	.420	.547	---	.485	1.427	.682
D	1.307	1.152	.485	.648	1.645	1.705
E	2.197	.986	1.427	1.645	.916	1.588
F	.475	.626	.682	1.705	1.588	---

These six clusters will serve as the criterion for the discriminant analysis. A look at the inter- and intra-cluster average distances suggests that there may be real problems in trying to discriminate between even these clusters on the basis of academic measures since they are so nearly alike.

V. Determination of Variables to be Included in the Analyses.

On the question of finding relevant variables for predicting success at M.I.T. some ten years of experience have isolated a set of pre-matriculation measures known to be useful. In those same ten years many other measures were studied but rejected as adding nothing new. The following list represents these variables to which we will add grades in the first semester at M.I.T. for use in predicting later success.

A. Secondary School grades in
   1. Plane Geometry
   2. Algebra
   3. Trigonometry
   4. Physics

B. Rank in graduating class

C. College Entrance Examination Board examinations
   1. Scholastic Aptitude Test: Verbal
   2. Scholastic Aptitude Test: Mathematic
   3. Chemistry or Physics Achievement
   4. Advanced Mathematics

The measures taken from secondary school records are known to have a non-linear relationship with freshman grades at M.I.T. For some years "step" functions has been employed in predicting freshman grade averages as far as the school record is concerned. For the purposes of this study the advantage of this experience enables us to use the "step" functions directly to derive a single variable for the school record. Otherwise, we would have to use average grades and a normalized rank variable.

2

Another measure available from records in the admissions office is a subjective rating of professional promise, which is called the "personal rating". The bases for the personal rating are endorsements from teachers, principal or guidance officer, the application form, and a record of a personal interview. While this rating is made independently of academic appraisals it is nevertheless correlated with achievement at M.I.T. Past studies have not shown it to have different information in the sense of adding to regression efficiency, but it is available for use here if it seems desirable to do so.

Among the variables studied in the past but found not to be useful for regression purposes is the categorization by type of secondary school of students entering directly from secondary school. It, too, is available for use.

Applicants to M.I.T. have one choice between two afternoon achievement tests. They may take either the chemistry or the physics test. Should there be any relevance to using the type of science test taken by a student, the type is also available.

For predicting success, then, we know the following measures will be relevant:

1. Summary of secondary school grades and class rank.
2. Scholastic Aptitude Test-Verbal
3. Scholastic Aptitude Test-Math
4. Physics or Chemistry achievement test
5. Advanced math achievement test
6. M.I.T. grade in first semester chemistry

7. M.I.T. grade in first semester physics
8. M.I.T. grade in first semester humanities
9. M.I.T. grade in first semester mathematics

To which we can add the following if desired

1. Personal rating of "professional promise"
2. Type of secondary school: public or private
3. Science test: physics or chemistry

For use in the discriminant analysis it is obviously valid to draw on the measures already planned upon for regression purposes. Hence, the first nine variables above will be used. Preliminary studies were made to determine if any discriminating information existed in the optional three. Should they be shown useful, they would be included.

Making use of the clusters of majors determined in the last chapter, it was possible to find for each cluster the average personal rating. The ratings are on a scale from 5 (poor) to 10 (superior) and have approximately a normal distribution. The following are the cluster mean personal ratings.

Major Cluster	No	Mean personal rating
Physics and Engineering	615	7.76
Chemistry and Engineering	293	7.86
Electrical Engineering	329	7.69
Physical Science	216	7.85
Chemical Science	102	7.83
Architecture	55	7.75

While these do not show wide differences, it was decided that the double potential in both the prediction and the discriminant analysis was sufficient to warrant the inclusion of the personal rating in the analyses.

Frequency counts for the six clusters against the public-private and physics-chemistry dichotomies were found as below when the data were known.

Major Cluster	Type of School		Science Test	
	Public	Private	Physics	Chem.
Physics and Engineering	480	139	303	161
Chemistry and Engineering	242	52	99	137
Electrical Engineering	284	45	195	62
Physical Science	176	40	113	60
Chemical Science	84	18	46	38
Architecture	41	14	28	9

The chi-square test of significance for the independence of the type of school versus the majors, gave an approximate probability value of 0.01 for randomness as the explanation of differences. The source of nearly half of the chi-square value was the fact that less private school graduates choose electrical engineering than would normally be expected. There seemed to be no rational explanation for this, so despite the apparent "significance", type of school was considered no further on the grounds that the single observation had no additional support empirically or logically.

By contrast, the chi-square for the physics-chemistry breakdown was beyond the range of most tables, a value of about 74 with 5 degrees of freedom. Further, the entries show logical preferences among the takers of physics for majors based upon physics and the complement of this for chemistry takers.

Since the type of test in science could also influence the reading of other measures, it was decided to make separate analyses for each group--takers of physics and takers of chemistry.

In the case of counseling M.I.T. freshmen in their second semester there is available other information regarding the choice of career. While the first official choice comes late in the freshman year, freshmen have given their preferences on the application for admission and at their first registration. At the beginning of this study it was planned to use these preferences in obtaining the probabilities of the various careers. At that stage it was not known exactly how mathematically these data would be treated. With students dividing their preferences among some twenty-five careers the denominators of proportions became small. Hence, the estimation of probabilities becomes quite inefficient. At this point the following argument was raised, though it does seem motivated by rationalization.

The statistic sought, namely the joint probability of membership in a career group and graduation, is a statistic which should summarize the relevant data in a form which enables comparison with non-objective considerations. The statistic should enable a freshman to combine the relevant data with his own interests and aspirations. This implies leaving factors involving interests and aspirations outside the basis of the statistic and suggests that the statistic be based solely upon objective measures. Making this clear to the student would still make the statistic meaningful to him. He would be freed of any preoccupation with numbers and can explore, with the aid of his adviser, his individual interests.

The above argument is tenable when the decision is to be made by the individual. In situations involving decisions by some institution the argument in invalid. In selecting students or in selecting employees, one would not dismiss so lightly measures related to interests. In the case at hand, however, we can justify the exclusion of measures related to interests provided the student or his adviser insists upon their inclusion at a later stage. The tentative choices of career fall into this category since they would be dictated largely by individual interests. So while these choices would surely enable us to make better predictions for ultimate career choices, they would best be excluded from our predictions. Our probability could then be given the following interpretation; "Based upon the academic records of your predecessors xx% of the students with records like yours chose to major in xxx." Should it turn out that academic records do not contain much information regarding career choices, this fact would be significant for the student in his decision making.

VI. Computations of Prediction and Discriminant Functions

The classes of 1956 and 1957 serve as the "normative" group from which we will estimate probabilities. The validity of these estimates will be tested on the class of 1958. Should this test show the estimates to be poor, it would mean the whole principal of estimating probabilities from past proportions is not valid for the measures employed.

As the basis for the estimation of the probabilities of success we will consider two criteria. The first and most conventional criterion is to regress four-year grade averages upon the measures employed. The second criterion calls for the single linear form of the measures that makes the largest value of "student's t" for the difference between means of those graduating and those failing to graduate. As described earlier, each of these analyses will be made for two groups--those who took the chemistry achievement and those who took the physics achievement.

Obviously the regression of four year averages upon the set of independent variables involves only those taking the normal four years to complete a four year program. For those in five year programs we used their five year average, but this involves only the architecture majors.

For ease in computations only the first two digits of the usual three involved in College Board scores were used. M.I.T. grades were treated on the scale where: A is 5, B is 4, C is 3, D is 2, E(conditional failure) is 1, F (complete failure) is 0. Table X gives the raw regression weights for the regression functions derived from each group.

Table X

Raw Regression Weights; Four Year Average as Criterian

Variable	Chemistry Group	Physics Group
Personal Rating	.502849	.171443
High School Record	.035198	.035465
Math Aptitude CEEB	.089846	.037557
Adv. Math Achievement CEEB	.025465	.005388
Science (Phy.-Chem) CEEB	.053955	.043958
Verbal Aptitude CEEB	.012172	.009760
M.I.T. Grade:Chemistry	1.694272	1.788739
M.I.T. grade:physics	1.133243	1.233699
M.I.T. grade:humanities	1.481118	.769297
M.I.T. grade:math	1.032742	1.114105
Number in Sample	316	493
Multiple Corr. (R)	.795	.754

To interpret the meaning of these functions in terms of the original set of variables we can find the correlation between the regression function and each of the independent variables. These are reported in Table XI.

Table XI

Correlations of Regression Function with Independent Variables

Variable	Chemistry Group	Physics Group
Personal Rating	.373	.280
H.S. Record	.549	.641
CEEB: Math Apt.	.306	.437
CEEB: Adv. Math Ach.	.458	.488
CEEB: Science (Phys-Chem)	.521	.448
CEEB: Verbal, Apt.	.358	.357
M.I.T. grade: chemistry	.847	.885
M.I.T. grade: physics	.814	.817
M.I.T. grade: humanities	.578	.499
M.I.T. grade: math	.772	.783
Number in Group	316	493

Despite the negative weights both functions are positively correlated with all the variables used. As we would expect, the functions are correlated more with M.I.T. grades than with other data, which demonstrates their relatively greater importance.

The other procedure is the case of using two groups as the criterion for a discriminant analysis. One group is composed of those graduating at the scheduled time depending upon their being in a four-year or five-year program. The remainder are those who fail to graduate as well as some who withdrew in good standing. A rather arbitrary rule was employed to "identify" the two other groups. Since reasons for withdrawal were unknown, it was assumed any withdrawn student with a C average or better at the time of withdrawal could be considered as "in good standing". Using this rule three groups were identified, but for our purposes we want to discriminate between graduates and non-graduates. We can let those withdrawing in good standing fall where they may.

The computation of the linear function is described in Rao (7) (pp. 287-289) but it involves rotation of a vector of differences between means of the two groups into a new space which is determined by the variances and covariances of the independent variables. Table XII gives the solutions for the two groups.

Table XII

Two Group Discriminant Function Weights; Graduates vs. Nongraduates as Criterion.

Variable	Chemistry Group	Physics Group
Personal Rating	.028863	.007817
H.S. Record	.002062	.005413
CEEB:Math Apt.	.007970	.041159
CEEB:Adv. Math. Ach.	.018538	.031692
CEEB:Science Phy.-Chem.	.004129	.001767
CEEB:Verbal Apt.	.020992	.079506
M.I.T. grade: chemistry	.172270	.923219
M.I.T. grade: physics	.100073	.327068
M.I.T. grade: humanities	.244576	.465042
M.I.T. grade: math	.133654	.279256
Number:Graduated	300	479
Number:Not Graduated	133	244
Number:Withdrawn	15	26
TOTAL	448	749

The correlations between these two discriminant functions and the original variables are given in Table XIII.

Table XIII

Correlations of Graduate-Nongraduates Discriminant Functions with Independent Variables

Variable	Chemistry Group	Physics Group
Personal Rating	.313	.216
H.S. Record	.500	.424
CEEB:Math Apt.	.239	.312
CEEB:Science (Phys-Chem)	.232	.354
CEEB:Verbal Apt.	.079	.025
M.I.T. Grade: Chem.	.814	.918
M.I.T. Grade: Physics	.757	.769
M.I.T. Grade: humanities	.603	.457
M.I.T. Grade: math	.792	.763
Number in Group	448	749

As before, all correlations are positive. However, the correlation between the functions and verbal aptitude is smaller than in the conventional regression system. Such a difference could be attributed to the fact that the discriminant approach included those not reaching the fourth year on time. Only after validation on the class of 1958 can we find which is the best basis for finding the probability of graduation.

Computation of Discriminant Functions

For purposes of computing the probabilities a given student will enter each of the six clusters of college majors, we make the assumption that the ten independent variables are normally distributed within each cluster and that each of the six density functions has the same variance-covariance matrix. Then we use the approach described in Chapter II to find the probabilities. Since there are six clusters, it is possible to use a transformation of the original ten variables into a space of five dimensions in which no information is lost. The use of discriminant techniques allow one to find the bases for the five dimensional space and arrange them in decreasing order of discriminating information. The basis vectors are solutions to the matrix equation

$$(A - \lambda W)\,V = 0 \tag{6.1}$$

where A is the matrix of sums of squares and cross-products of the group centroids in the ten dimensional space, W is the pooled sums of squares and cross-products within the group, V is the latent vector which determines one of the new basis dimensions, and λ is the latent root associated with that vector. The value of λ associated with each vector solution enables the ranking of the new basis vectors since each λ is proportional to the F ratio (for that basis vector) of the analysis of variance of the group means. If there should actually be only one true dimension which discriminates, its λ value will be large and the other λ values will be only randomly different from zero. In more intuitive terms, the λ values help us to separate what we think to be "noise" from what we think to be "signal". Should we decide, say, that of five λ values two are only randomly different from zero, we could hope that all relevant information is in the other three dimensions. Until we actually validate such analyses we don't know what is actually relevant. Tests of significance would be of some help here, but at present there are no good tests for this purpose that are less subjective than the above. The cross-validation will serve as the best test we could make.

The matrix equation given in (6.1) was solved for the chemistry and for the physics group. The values of λ for each set are given in Table XIV in decreasing order.

Table XIV

Values of λ (Latent Roots) in Solutions to Matrix Equation 6.1

λ	Chemistry Group	Physics Group
1st	.4056	.1823
2nd	.3453	.1674
3rd	.0717	.0447
4th	.0384	.0231
5th	.0177	.0053

Using the Bartlett approximation discussed by Rao (7) on pp. 370ff, tests were made to see if the last of each set of these λ values could be only randomly different from zero. Then another test was made to see if the last two values (corresponding to the two basis dimensions with the least discriminating information) could be randomly different from zero, and so on. This series of tests is summarized in Table XV.

Table XV

Approximate Tests of Significance of Discriminant Functions

Test	Chemistry			Physics		
	Chi-Sq.	d.f.	"P"	Chi-Sq.	d.f.	"P"
5th zero	1.648	6	.75-.90	5.143	6	.50-.75
4th & 5th zero	5.134	14	.98-.99	27.158	14	.01-.02
3, 4, 5 zero	66.256	24	0-.001	68.250	24	0-.001
2, 3, 4, 5, zero	265.088	36	0-.001	197.898	36	0-.001

The ranges of "P" are for the probability that apparent discrimination in the corresponding dimensions could be attributed to random factors alone. In the chemistry group it appears the discrimination in the fourth and fifth dimensions is only randomly different from zero. In the physics group it is questionable whether the fourth dimension could be attributed to chance or not. Similar experiences have shown that quite often such a questionable dimension is only "noise". For this reason it was decided to employ the three dimensional space for each group. The weights corresponding to these three linear forms for each group are given in Tables XVI and XVII. In each instance they have been normalized to be direction cosines for the new basis vector.

Table XVI

Weights for Three Basis Vectors of Reduced Space: Discrimination of Clusters of Majors; Chemistry Group

	Direction Cosines		
Variable	First	Second	Third
Personal Rating	-.127860	.113968	-.055106
High School Record	.094084	.037719	-.311913
CEEB:Math Aptitude	.034854	-.026724	-.139738
CEEB: Adv. Math. Ach.	.019546	-.030998	.102610
CEEB: Science (Chem.)	-.003794	.042869	.041206
CEEB: Verbal Aptitude	.059383	.021594	.001306
M.I.T. Grade: Chemistry	-.646721	.799768	-.064827
M.I.T. Grade: Physics	.364251	-.511879	.449046
M.I.T. Grade: Humanities	-.020352	.270024	.724778
M.I.T. Grade: Math	.646778	-.083954	-.369824
Standard Deviation	1.118819	.812306	1.261341
Means:			
Cluster A	7.0860	2.8136	-.0855
Cluster B	6.6788	3.2700	-.0340
Cluster C	7.5126	2.9170	-.2207
Cluster D	7.9642	3.4246	.1953
Cluster E	6.9542	3.7099	-.3810
Cluster F	6.1536	3.0011	1.2882

Table XVII

Weights for Three Basis Vectors of Reduced Space: Discrimination of Clusters of Majors; Physics Group

	Direction Cosines		
Variable	First	Second	Third
Personal Rating	-.326588	.000216	.110586
High School Record	.090519	.045028	-.068555
CEEB: Math Aptitude	.023709	-.037617	.054146
CEEB: Adv. Math. Ach.	-.012673	.049011	.193944
CEEB: Science (Phys.)	.084501	.010875	-.140969
CEEB: Verbal Aptitude	-.018844	.042297	.045296
M.I.T. Grade: Chemistry	-.670572	.885929	-.647109
M.I.T. Grade: Physics	.599878	-.405224	.399459
M.I.T. Grade: Humanities	.024578	-.070920	.584708
M.I.T. Grade: Math	.258463	-.195297	.019981
Standard Deviation	1.041452	.858386	1.523320
Means:			
Cluster A	3.3185	4.5034	11.4617
Cluster B	3.0091	5.1854	11.1863
Cluster C	3.8988	4.7114	11.1345
Cluster D	3.9502	5.1114	11.8755
Cluster E	3.1444	5.3703	11.1871
Cluster F	2.6553	4.6542	12.1352

On the assumption that the six clusters are normally distributed in the original ten dimensions, they are also normally distributed on the three new dimensions. In addition to this the new dimensions are uncorrelated, which means the computation of a density in the three dimensions is the product of the three marginal densities. Hence, to compute for a person each of the probabilities of his being in a particular cluster of majors we can follow the following steps.

1. Find from his ten original scores his values on the three new dimensions. Call them W_1, W_2, W_3.

2. Find for each of the six groups the sets of three deviations from the cluster means on the new dimensions. For example, for cluster B, we find

$$\begin{aligned} &W_1 - \mu_{1B} \\ &W_2 - \mu_{2B} \\ &W_3 - \mu_{3B} \end{aligned}$$

where the μ's are the means for cluster B.

3. Divide each deviation by the corresponding standard deviation. For example,

$$\frac{W_1 - \mu_{1B}}{\sigma_1} ; \quad \frac{W_2 - \mu_{2B}}{\sigma_2} ; \quad \frac{W_3 - \mu_{3B}}{\sigma_3} .$$

4. Find one-half of the sums of the squares of the values just obtained.

$$\frac{1}{2}\left(\frac{W_1-\mu_{1B}}{\sigma_1}\right)^2 + \frac{1}{2}\left(\frac{W_2-\mu_{2B}}{\sigma_2}\right)^2 + \frac{1}{2}\left(\frac{W_3-\mu_{3B}}{\sigma_3}\right)^2$$

(Note: By adjusting the weights used to find the W's by dividing each weight by $\sqrt{2}\sigma_i$; and each mean by the same value, this step can be achieved by finding

$$(W_1'-\mu_{1B}')^2 + (W_2'-\mu_{2B}')^2 + (W_3'-\mu_{3B}')^2$$

where the primes designate the adjusted W's and μ's.)

5. Find the exponential of the values obtained. This gives a value for each cluster.

6. Find the quotient of the number in the normative group divided by the exponential for that group.

7. Find the sum of these six values.

8. Find the quotient of each of the six values divided by the sum. These are the proportions desired.

The steps above could have been completed with some revisions with the original ten dimensions. The procedures would have been longer, however. Using the shorter method involving only three dimensions does not sacrifice any discriminating information provided the original assumption of normal distributions is valid. Actually, the assumption can be somewhat invalid and still the reduced space contains whatever information is valid.

As described above the probabilities were computed for all students in the study, both normative and test groups. Also, the two functions being considered as the basis for computing the conditional probabilities of success were computed for both groups. The results of these computations are discussed in the next chapter.

VII. Tests with a New Group of Students - Class of 1958

Validation of Discrimination Between Clusters of College Majors.

Using the procedures outlined in the previous chapter estimates were made of the probability a student would choose a major from cluster A, cluster B, etc. For each student in both normative and test groups there were six such probabilities computed. To measure the efficiency of these probabilities we can determine how often it "makes mistakes" by the following logic.

If a student actually chooses a major in cluster A, a perfect description for him would have the probability of A being unity and the other probabilities being zero. For the student who actually chooses another cluster the "mistake" for him with respect to cluster A is his probability of cluster A. We can therefore evaluate the efficiency of predicting membership in cluster A by adding all the $\left[1-P_i(A)\right]$ terms for those choosing A and all the $P_i(A)$ for those not choosing A. Similarly we can do this for each cluster. Doing this is comparable to finding a biserial correlation coefficient but has the advantage of a physical interpretation. When we average the sums above we get an average deviation between the perfect probability and the computed probability.

The probabilities we have computed are ex posteriori probabilities in which the data is used to improve upon the a priori probability, which is just the proportion of the total in the given cluster. Should the ex posteriori probabilities really improve the a priori probabilities, we should observe a reduction in the average deviations discussed above. We can therefore compare the averages for the two kinds of probabilities: 1) one kind which is constant for all students; 2) one which varies according to scores on the variables used. The averages are compared below in Table XVIII.

TABLE XVIII

Averages of Difference Between Probability of Given Cluster and "Perfect" Probability for Each Cluster

	1956 & 1957 Classes (1197 Students)		1958 Class (699 students)	
Cluster	A Priori Probabilities	Ex Posteriori Probabilities	A Priori Probabilities	Ex Posteriori Probabilities
A	.468	.428	.405	.435
B	.306	.272	.322	.294
C	.329	.306	.333	.319
D	.240	.222	.250	.231
E	.125	.118	.129	.121
F	.047	.046	.042	.039

In each instance the averages above show substantial differences between the probabilities and the perfect case. While the *ex posteriori* probabilities do show some reduction in the differences in the classes of 1956 and 1957, the differences are not very large. In the class of 1958 the differences are even less.

The meaning of these results statistically is that the use of the academic measures does not improve to any great extent the *a priori* probability. We can use the *a priori* probability and ignore the academic data and do essentially as well as we can by using it to derive a probability.

In the counseling situation these results are quite significant as far as choosing a career at M.I.T. is concerned. They indicate that neither a student nor his adviser should be occupied with his profile on the academic measures studied here as far as judging his similarity to those in various majors is concerned. Both parties should rather be concerned with other kinds of information, notably the student's interests and ambitions.

It must be noted that these results and their interpretation are valid only at M.I.T. and should not be translated into other contexts. Further, these results should not be interpreted to mean that the theoretical bases for determining the probabilities is suspect. The only inference that can truly be drawn is that at M.I.T. variations among students in terms of the academic measures studied are far in excess of the variations that can be discerned between the various college majors.

In terms of our study, the original objective of finding from any individual's profile of academic scores the probability of his choosing any of several careers has failed to improve upon the trivial *a priori* probability. While this has implications for future counseling of students, it means part of the proposed statistic is of no consequence at M.I.T.

Testing of Functions for Use in Computing Probabilities

The next validation we are concerned with involves the question of which of two linear forms is best as the basis for estimating the probability of graduation. We are considering either a regression function which uses the four year average as its criterion or a two group discriminant function which uses all graduates and all non-graduates (including those not reaching the fourth year) in its derivation. We can test these out in the group from which they are derived (classes of 1956 and 1957) as well as the test group, the Class of 1958.

As a means for evaluating the different functions we can use the normalized difference between the means of the graduated and not graduated groups. This is, of course, the criterion for one of the functions, and we should expect in the normative group that that function would be better to some degree. Our real concern is whether this function retains its superiority in the test group. Should that be the case, it means that that function involves some meaningful factor or factors missing from the other function. On the other hand, an indication that its superiority is not retained means the apparent superiority was not real. Following are the results of the testing of the two functions. Those withdrawing in good standing are excluded.

TABLE XIX

Student's "t" Test Between Graduates and Non-graduates for Two Functions with Different Criteria
Chemistry Group

	Normative: '56 and '57	Test: '58
Differences: Regression Function	10.56	8.83
Difference: Discriminant Function	11.55	8.07
Number graduated	300	164
Number not graduated	133	71

Table XIX (Cont.)

Physics Group

	Normative: '56 and '57	Test: '58
Difference: Regression Function	11.85	9.33
Difference: Discriminant Function	13.20	9.58
Number Graduated	479	305
Number not Graduated	244	114

The superiority in the normative group is not marked. In the test group there is no reason to choose one function over the other. This means the analysis based upon the larger group of students shows no improvement. Hence, the more conventional system is as good as the one proposed as an improvement. In fact, a combination of the two independent values which squares them and adds them shows the conventional system to be superior in the test group, but not significantly so.

Given that we want to find probabilities of graduation from the regression function the question arises as to the mathematical means for doing so. If graduation were determined solely from the four-year average such that those above a specific average would be graduated and those below not, we could use the conventional expectancy table. This employs the specific cutoff and the standard error of estimate. There is no such cutoff on four-year averages, however. Instead graduation is determined by a sequence of several decisions. Hypothetically some scale does exist from which we could evaluate a cutoff and a standard error of estimate if we knew where people fell on the scale. Then we would employ the table of areas under the normal curve to create a functional relationship between probabilities of graduation and the regression function. This gives an S-shaped curve called the ogive. The logistic curve is another S-shaped curve involving two parameters. Either ogive has two parameters; one determines the point on the scale of the independent variable corresponding to a probability of one-half, and the other determines the slope of the ogive at that (or any other) point.

Our job here is to find an ogive which would enable us to easily adjust the parameters for each department so as to reflect differences between them in difficulty of achieving graduation. Since the hypothetical scale is not available to us, the approach using an expectancy table seems unfruitful except by approximations. Exactly how to make differences between college majors influence these approximations is not known. There is, however, a basis for the logistic curve which will enable us to make approximations in a rather crude way. In other studies, however, this crude approach has given quite satisfactory results. If the regression function values for graduated and non-graduated groups are normally distributed with the same variance, use of Bayes' Rule as in equation (1.4) gives a logistic curve as follows.

$$(7.1)\quad \text{prob (graduation}\,|\,x) = \frac{1}{1 + \frac{N_{NOT}}{N_{GRAD}} \exp\left\{\frac{\mu_{NOT} - \mu_{GRAD}}{\sigma^2}\left(x - \frac{\mu_{GRAD} + \mu_{NOT}}{2}\right)\right\}}$$

where N_{NOT} and N_{GRAD} are the numbers not graduated and graduated, μ_{NOT} and μ_{GRAD} are the means of x values for each group, and σ^2 is the common variance. By manipulation this can be put into the form of the logistic function

$$(7.2)\quad \text{Logistic}\,(x) = \frac{1}{1 + e^{-a(x-b)}}$$

where

$$(7.3)\quad a = \frac{\mu_{GRAD} - \mu_{NOT}}{\sigma^2}$$

and

$$(7.4)\quad b = \frac{\mu_{GRAD} + \mu_{NOT}}{2} + \frac{\sigma^2}{\mu_{GRAD} - \mu_{NOT}} \quad \log_e \frac{N_{NOT}}{N_{GRAD}}$$

The "a" value determines the shape of the logistic curve at any point. The larger "a" is the steeper the logistic is at any point x. When x equals b, the value of the logistic is one-half. For two groups of college majors whose "b" values are identical, a change in the "a" value alters logistic values very little. For two groups of college majors whose "a" values are identical, a change in the "b" values does alter the logistic values by moving the logistic curve itself to the right or left. This movement from side to side reflects the first order of differences in difficulty of the various majors. Hence, we could allow the "b" values to differ from group to group, keep the "a" values constant, and achieve a reasonable estimate of the differences between the groups.

Such procedures are admittedly crude. Several complicated means are available for fitting logistic curves which are mathematically much more sound (see Silverstone, 8). Two reasons prompted us to employ the method used. First, it is simple to use and experience has shown it gives a reasonable fit. Second, the opportunity to make differences easily between the "b" parameters for each group is a distinct advantage. By validation procedures we can judge whether these differences actually increase the efficiency of our probability estimates.

One very real problem we have involves the efficiency of estimating the "population" values of the parameters that determine the logistic curve. To estimate the parameters for a specific major we can only use whatever number we have in that major in the classes of 1956 and 1957. When this number is small clearly we have poor estimation. While it is true that we will also have a small number in the succeeding class and can therefore expect the poorer estimates to affect only a few, we would prefer not to be considerably off in estimating for even a few students. Hence, we must weigh a more comfortable conservatism against the potential sacrifice of valid information obscured by small numbers.

The ratio

$$\frac{N_{not}}{N_{grad}}$$

is actually a ratio of the a priori probability of not graduating to the a priori probability of graduating. Clearly this ratio is influenced by small numbers, and variations between the various majors in this ratio could be quite random. The conservative choice is to use the same ratio for all students regardless of major. While this would remove certain differences between the majors, we would be less inclined to go wrong when dealing with the smaller majors.

The ratio $$\frac{\mu_{GRAD} - \mu_{NOT}}{\sigma^2}$$

determines the slope of the logistic curve. While we argued above that differences between the majors in the true slope are of second order, we also have some confidence that the difference in the numerator is fairly constant from major to major. Conservatively using the same value of this ratio for every department is not expected to sacrifice much information.

The ratio $$\frac{\mu_{GRAD} + \mu_{NOT}}{2}$$

is left. It too is subject to variations of a random sort due to small numbers. It is, however, more stable than the other ratios. Further, it determines the 50-50 point on the logistic curve, which we have considered as the first order difference between the majors. Hence we will use a different value of this ratio for each department.

All the parameters needed were computed from the 1956 and 1957 classes again keeping the physics and chemistry groups separate. The values of the logistic were computed for each student in the class of 1958 according to his final choice of major. As was done with the probabilities of the different clusters of majors, we computed the average difference between the "perfect" probability and a probability computed under each of three conditions:

1. The same *a priori* probability assigned to all students regardless of major or scores on academic measures.
2. A logistic-type probability computed from the four year regression function but ignoring choice of major.
3. A logistic-type probability computed from the four year regression function using the choice of major.

The results of these computations are summarized in Table XX.

TABLE XX

Average Difference Between "Perfect" Probability of Graduation and Computed Probability; Class of 1958

Probability	Chemistry Group	Physics Group
1. *A priori*	.424	.417
2. Logistic-type without using choice of major	.316	.352
3. Logistic-type using choice of major	.309	.350

While it is true a substantial reduction in these average errors is achieved by using the academic data, there seems little use to employ also the choice of major. One could challenge the adequacy of the logistic curve, but its use did allow a simple distinction to be made between majors. In addition this distinction allowed for a more obvious difference between the majors.

As a demonstration of how adequate the logistic curve is as a means for finding the S-shaped curve below are given some tabulations of the class of 1958. At different levels of the four year regression function are given in Table XXI the number graduated, the number not graduated, the proportion at that level graduated. Note that the scale of the regression function is completely arbitrary.

TABLE XXI

Distribution of Regression Scores (Arbitrary Scale) for Class of 1958

Graduates, Non Graduates; Proportion in Each Interval Who Were Graduated

Chemistry Group

Regression Function Interval	Number Graduated	Number Not Graduated	Proportion Graduated
120-139	0	1	.000
140-159	0	2	.000
160-179	2	6	.250
180-199	1	7	.125
200-219	4	9	.308
220-239	7	11	.389
240-259	17	14	.548
260-279	21	7	.750
280-299	23	4	.852
300-319	29	3	.906
320-339	15	2	.882
340-359	16	3	.842
360-379	10	2	.833
380-399	11	0	1.000
400-419	7	0	1.000
420-439	1	0	1.000

TABLE XXII

Distribution of Regression Scores (Arbitrary Scale) for Class of 1958 Graduates, Non Graduates; Proportion in Each Interval Who are Graduated

Physics Group

Regression Function Interval	Number Graduated	Number Not Graduated	Proportion Graduated
90-109	1	2	.333
110-129	0	0	----
130-149	1	1	.500
150-169	3	8	.273
170-189	2	6	.250
190-209	3	7	.300
210-229	10	15	.400
230-249	7	6	.538
250-269	10	10	.500
270-289	27	14	.654
290-309	46	12	.793
310-329	27	12	.692
330-349	30	6	.833
350-369	35	4	.890
370-389	20	7	.741
390-409	31	1	.968
410-429	17	2	.894
430-449	17	1	.944
450-469	13	0	1.000
470-489	5	0	1.000

The preceding tables are given in graphic form in Figures 2 and 3. In each graph the proportions are plotted. Also in each graph are given the logistic curves derived from the corresponding 1956 and 1957 group. The adequacy of the fit attests first to the relevance of the logistic curve as derived and further to the means by which the parameters are estimated from one group for use with another. It must be remembered that the curves are estimated from the normative group and the empirical points are from the test group.

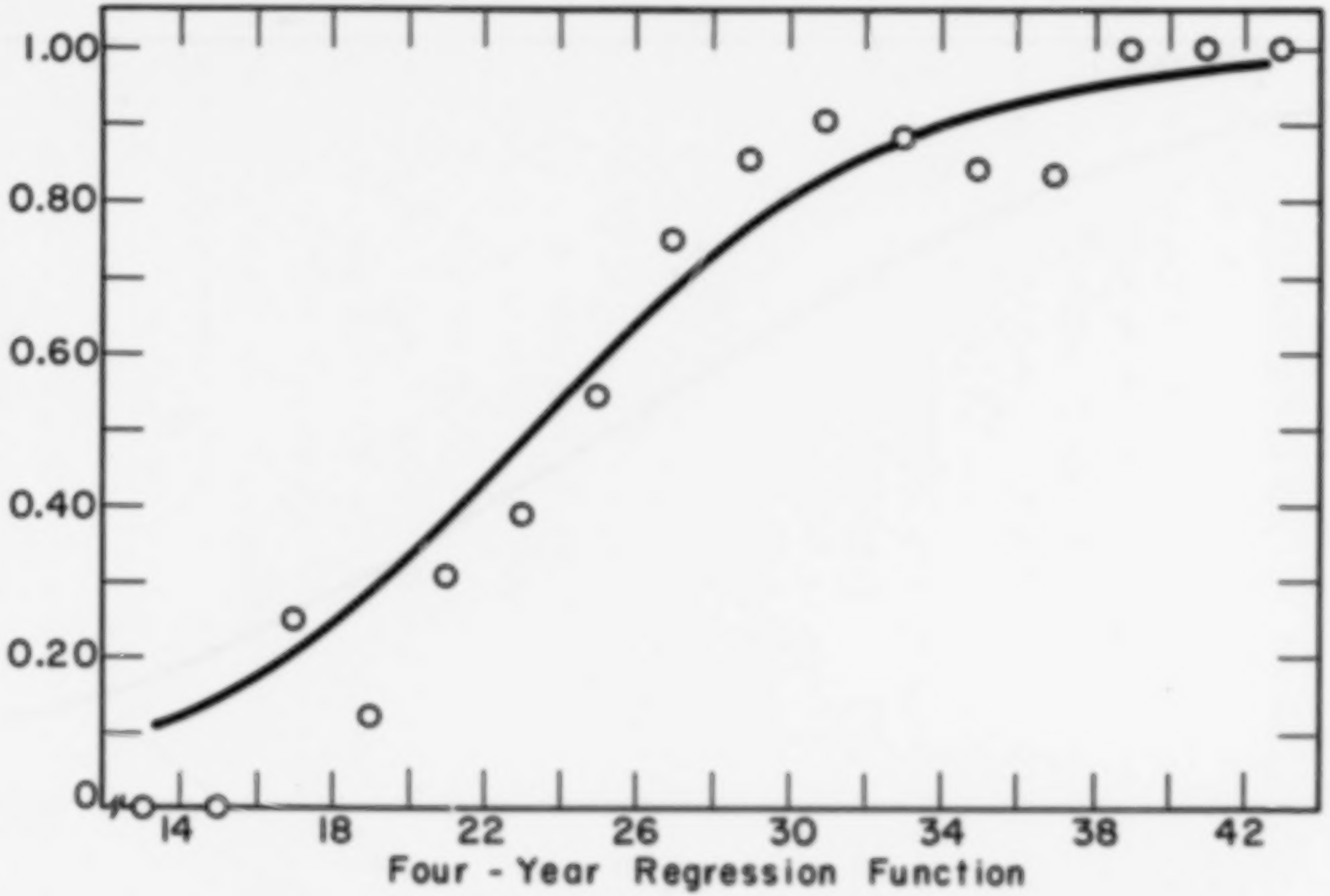

FIG. 2 COMPARISON OF LOGISTIC CURVE AND PROPORTIONS GRADUATING. CLASS OF 1958: CHEMISTRY GROUP

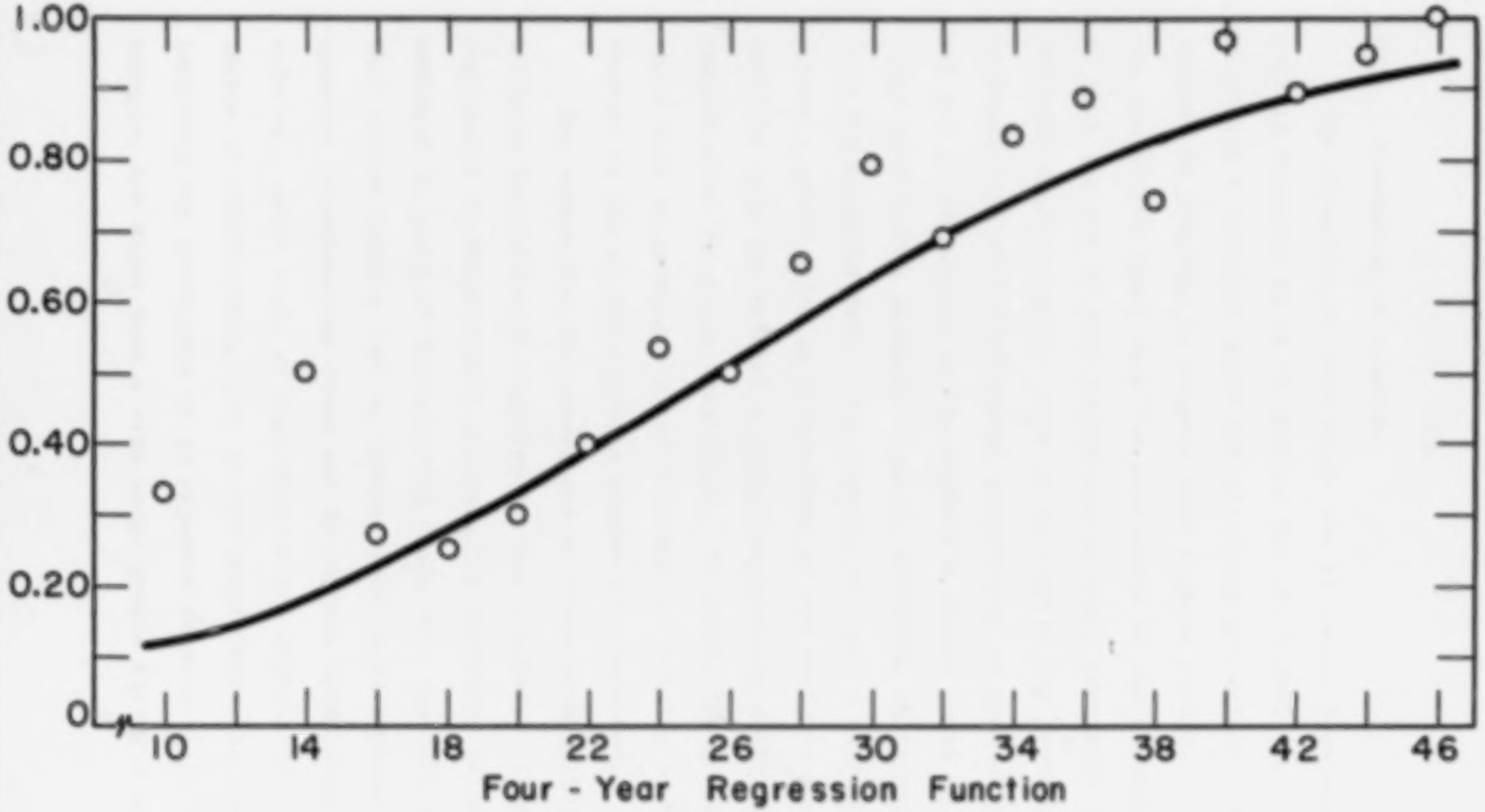

FIG. 3 COMPARISON OF LOGISTIC CURVE AND PROPORTIONS GRADUATING. CLASS OF 1958 : PHYSICS GROUP

VIII. Discussion of Results

The objective of this study was to employ techniques of profile analysis so as to develop the joint probability of selecting a specific major and attaining graduation. For any student we proposed to compute such a joint probability at the end of the first term for each major he was considering so that the set of such probabilities would summarize all relevant academic data. Were we to compute the probabilities we would find each conditional probability of graduation given any one of the majors as the student's choice would equal each other conditional probably--clearly providing no information from the academic data. The probability that a student will choose a given major is independent of his academic profile and would be only the trivial *a priori* probability which is quite removed from the given individual. In short, the probabilities would fail in passing new and relevant information to the student as far as his academic scores are concerned.

One reason for the inadequacy of these probabilities involves the criterion employed to test adequacy. In either regression or discriminant analysis the customary criterion employed in tests of significance deals with the null hypothesis that random factors just by chance alone could explain the apparent relationships drawn out by certain optimizing procedures. While tests of significance are associated with the degree of relationship, they do not answer directly the question involving the usefulness of an apparent relationship. For example, one might have a very large sample giving a multiple correlation of 0.2 which would be judged significantly different from zero. This would provide, however, very little confidence in a prediction system. The criterion employed here dealt with the adequacy of a probability computed for each individual when compared to the true probability which takes on only the values of zero or one. If the average deviation between the computed probability and the true probability could be reduced as much as 0.1 by the use of certain measures, then the computed probability based upon these measures would seem to be useful. That is, improvements affecting the first decimal of the probability might be considered worthwhile. In the section of this study dealing with the probabilities of choosing the various majors the use of the measures barely improved upon the second decimal of the *a priori* probabilities. In the section dealing with the probability of graduation, the use of the measures did achieve an improvement over the *a priori* probability that affected the first decimal. The effort to compute the probability of graduation conditional upon the choice of major showed improvement over the unconditional probability that barely affected the second decimal.

The import of these findings is not that differences do not exist between these various majors on the measures studied, rather it is that efforts to employ these differences to compute probabilities for use in the decision process show such use to be questionable. The "noise" inherent in such computations is almost of the same magnitude as the differences used. While this sounds like a study that failed, it actually has considerable significance. It means that observed differences between college majors on the measures studied have less relevance to the choice of a student than would be expected. Were we to use a complete decision-making model as developed by statistical decision theorists (see, for example, Luce and Raiffa, 4) the measures would have little effect on the decision.

The full import of this study takes several directions. There are often inclinations to attach significance to the interaction between different measures in the context of assessing an individual's record. We might, for example, be inclined to tell a student with lower grades in chemistry than in his other subjects that he should stay away from a major involving much chemistry in the curriculum. Empirically, however, there is no evidence that students have done this uniformly in the past. We might also be inclined to tell the student that his chances of success in a chemistry-oriented major would be poorer than his chances in other majors, but this cannot be validated empirically either. An apparent reason for the absence of such validations is the fact that chemistry grades measure an academic endeavor which differs in name from other endeavors but which is nevertheless an academic endeavor. The implications of poor grades in chemistry are more relevant to the general academic achievement than their differential relevance to achievement in just chemistry-oriented endeavors.

An extreme example of the general validity of some measures in contrast to their differential validity involves work in the humanities. A common hypothesis regarding the prognosis of M.I.T. student records leads to the emphasis of measures of achievement in scientific fields as more important than measures of achievement in humanities. It seems reasonable to postulate achievement in science as a necessary condition for success at M.I.T. (which it is) but to consider achievement in humanities as desirable but not necessarily necessary. The statistical facts belie this for good reason. Measures of scientific achievement are often highly correlated among themselves. Further, they are less correlated with measures of non-scientific achievement. While the non-scientific measures are often less directly related to success at M.I.T. than the scientific measures, they have importance due to relevant factors they measure uniquely. In short, one can make a better prognosis of success at M.I.T. knowing a science grade and an humanities grade than he can knowing two science grades.

The fact that this study has shown the measures studied to be of little relevance to choosing a career once a student registers at M.I.T. raises the question of what should be done to find measures that are relevant. One plausible explanation is that the first term involves too many problems of adjustment and transition. The weakness for prognosis purposes that we discovered may not be the case were we to include more than the first term grades. A separate study might be made for this purpose, but that study would not involve the counseling situation considered here. One might propose that a more clinical and less actuarial approach would develop relevance still in the area of assessing intellective factors. The book by Meehl (5) contains evidence that the clinician is actually less efficient at assessing intellective factors than is the statistician. Then it might seem reasonable to assess non-intellective factors along the lines of the work described by Stern, Stein, and Bloom (9). One problem with the use of projective instruments for measuring personality characteristics deals with the fact that they can be "rigged" by an intelligent student who feels he wants for any of several reasons to pose himself different from what he is. Other problems more serious than these arise dealing with ethical and value considerations. It is open to question whether an institution should conduct analyses of its students in the dimensions of personality before it has complete assurance that doing so is in the best interests of the student. It can be questioned whether anyone beside the student should pick apart his unconscious for purposes of choosing careers. It can also be questioned whether personality characteristics should interact with career choices at all. Such problems are difficult to answer, but they are important considerations to an institution desiring to employ newer techniques in the counseling situation.

Bibliography

1. Cronbach, L. J., and Gleser, G. C., Psychological Tests and Personnel Decisions, University of Illinois Press, Urbana, 1957.
2. Fisher, R. A., "Inverse Probability," Proceedings of the Cambridge Philosophical Society, 26 (1930), 528-535.
3. Kempthorne, Oscar, The Design and Analysis of Experiments, John Wiley & Sons, Inc., New York, 1952.
4. Luce, R. D., and Raiffa, H., Games and Decisions, New York, John Wiley & Sons, Inc., 1957.
5. Meehl, Paul E., Clinical Versus Statistical Predictions, University of Minnesota Press, 1954.
6. Mood, A. M., Introduction to the Theory of Statistics, New York, McGraw-Hill, 1950.
7. Rao, C. R., Advanced Statistical Methods in Biometric Research, John Wiley & Sons, Inc., New York, 1952.
8. Silverstone, H., "Estimating the Logistic Curve," Journal of the American Statistical Association, 52 (1957), 567-577.
9. Stern, G. G., Stein, M. I., and Bloom, B. S., Methods in Personality Assessment, The Free Press, Glencoe, Illinois, 1956.
10. Tatsuoka, M. M., "Joint Probability of Membership and Success in a Group: An Index which Combines the Information from Discriminant and Regression Analyses as Applied to the Guidance Problem," Unpublished Doctor's Thesis, Harvard University, Graduate School of Education, Cambridge, Mass., 1956, 150 p.
11. Tiedeman, D. V., and Sternberg, Jack J., "Information Appropriate for Curriculum Guidance," Harvard Educational Review, 22 (Fall, 1952), No. 4, 257-274.